U.S. Department
of Transportation

Research and
Special Programs
Administration

John A. Volpe
National Transportation
Systems Center

Understanding and Predicting Traveler Response to Information: A Literature Review

Jane Lappin
John A. Volpe National Transportation Systems Center

Jon Bottom
Charles River Associates

Prepared for

Brian Gardner
U.S. Department of Transportation
Federal Highway Administration
Office of Metropolitan Planning and Programs
Washington, D.C.

December 2001

Table of Contents

Table of Extended Abstracts

Table of Extended Abstracts (continued)

Table of Extended Abstracts (continued)

Table of Extended Abstracts (continued)

Table of Extended Abstracts (continued)

1 BACKGROUND

1.1 Introduction

In the early days of automobiles when, for the first time in history, large numbers of people had opportunities to travel well beyond their local areas, finding directions was a problem. Prior to that, the range of most peoples' travels was limited to a relatively short distance from their home, and people quickly became familiar with the small network that they regularly used. Signage was not needed. However, as new drivers roamed into unfamiliar areas, the lack of signage made getting lost a common occurrence.

Technology soon provided solutions to this problem (French 1986). For example, it was possible to buy and install an in-vehicle cylindrical or disc-shaped device that advanced at a rate that was synchronized with wheel rotation. The cylinders or discs had way-finding information printed on them. When initialized with the correct trip starting location, information about the direction options at each major decision point would be displayed prior to arriving there. Some enhanced versions also included static travel information such as road conditions, and locations of unimproved railroad crossings and speed traps.

Over time, of course, major investments in signage and road maps made such devices less useful, and research in traveler information systems was limited to relatively specialized applications such as for military vehicles. It is only relatively recently, with traffic congestion and the externalities of automobile use becoming more of a concern, that "advanced" traveler information systems (ATIS), have again become of interest. Technological progress in vehicle location, traffic monitoring, and data processing and communications have made possible applications that were probably not imaginable in the early days of the field.

Travel-related messages may be derived from static or dynamic information about the network. Static messages provide fixed information about the network and the destinations that it serves, and may be of use in tasks such as way-finding or preliminary trip planning; however, they do not recognize actual traffic conditions and so cannot respond to them. Dynamic messages reflect either prevailing or predicted conditions on the network, and require capabilities for collecting and possibly processing network data in real time. Such messages may describe the network conditions, or make recommendations based on the conditions, or both. A variety of presentation media (graphical, spoken or text) and levels of quantitative or qualitative detail in the messages are possible.

With these recent enhancements of traveler information system technological capabilities has also come an increased interest in understanding how travelers react to information provided in this way. There are many reasons for this interest:

- public and private organizations developing travel information products need to know what product features are most valued by travelers and why; this knowledge enables better products to be designed, and appropriate pricing strategies to be elaborated;

- public agencies investing in travel information infrastructure also need to know how travelers perceive and value the benefits that they will derive from the provided information, as guidance in making economically worthwhile investment decisions;

- many of these same agencies are examining the contribution that traveler information systems may make towards improving the overall operation and performance of their transportation systems, either by themselves or in combination with advanced traffic management systems (ATMS). Such network-level ATIS impacts can best be determined by aggregating the individual responses of many travelers to the information that they are provided by ATIS, but in doing this the interactions of the travelers on the network may become important and then must also be taken into account;

- finally, ATIS technologies currently under development will eventually be able to provide information based on predictions of future travel conditions; however, such information must incorporate a forecast of how travelers will react to it. For example, on the basis of short-term traffic forecasts, an ATIS may inform drivers that a certain route is expected to become congested in the next hour. If drivers react to this information by choosing a different route, their response may invalidate the forecast, leaving the original route free flowing but creating even worse congestion on an alternate route. Generating guidance based on forecast traffic conditions requires being able to forecast how drivers will respond to the guidance that they receive, determining the aggregate network-level impacts of the responses, and incorporating those responses and impacts into the guidance itself.

1.2 Scope and purpose of this document

In view of these reasons for an interest in traveler response to information, the Federal Highway Administration commissioned a review of published information on the subject; this report is one of the products of the study. It is a review of the literature published as of mid-2001 on the topic of traveler response to real-time information at the individual and network levels. (Static travel information is only considered in passing because of its rather limited scope for improving individual decisions or affecting network conditions.) The report's intent is to summarize what is currently known about traveler response to information, in a form that provides a useful high-level understanding of the main issues.

This is not a comprehensive review – it could not possibly be, given the volume of material that has been (and continues to be) published in relevant areas. Several criteria were applied in deciding what to review:

- recent (past few years) publications with relevant research or applications results;
- publications providing summaries of long-term research or operational programs;
- selected early (pre-1990) publications, chosen for their historical interest or because their results are still relevant;
- selected publications from the mid-1990s, again chosen for their relevance or historical interest.

It will be seen that, despite the number of publications in the field, understanding of traveler response to ATIS is still in its initial stages. No one is yet able to accurately predict, for a VMS displaying a particular message at a particular location in a particular network, what the effect on individual travelers or on overall network conditions will be. Only limited data is available on individual responses to information, from operational deployments or from surveys investigating user reactions to hypothetical systems. Available data tends to be concentrated in specific areas such as commuter driving behavior; much less is known about information effects on non-commute trips, transit riders and commercial vehicle operators, for example. Efforts to develop models of traveler response based on these data are, for the most part, cutting-edge academic research far removed from the capabilities and needs of mainstream practitioners. Network-level forecasting models capable of predicting ATIS system impacts are also still mostly *ad hoc* in nature, frequently involving the cobbling together of two different model systems.

This state of affairs is not entirely surprising. Automobiles and modern transit systems were in use for roughly half a century before systematic and comprehensive travel data collection efforts were undertaken, and useful individual- and network-level transportation planning models began to be developed and routinely applied. While the pace of research and development is much faster now, a decade of experiments with ATIS is not foundation enough to support the development of a full understanding of its effects.

For these reasons, this review does not devote excessive effort to documenting the complete sets of results from available user surveys, or the full details of current model systems. For the same reasons, too, it discusses survey and analysis methods as well as with results, because robust and powerful methods will be needed to obtain further useful results in the future. At this point in the development of the field, the creation of appropriate tools and methods is just as necessary and important as their application.

This document may perhaps best be regarded as a source of raw materials that can be used in many different ways. Material can be extracted from it to prepare more specialized documents, focused on particular topics or audiences. It provides extensive references to and discussions of the published literature, enabling the original detailed results on particular subjects to be easily located. Although it mostly highlights what has been done to date, this focus also illuminates some of the gaps in current knowledge, and suggests actions that need to be taken in the future to advance the state of knowledge. In one particular area – the modeling of network-level ATIS impacts – the report makes suggestions regarding specific directions for future development approaches.

A companion document provides a number of specific recommendations for Department of Transportation actions to further knowledge in the field of traveler response to information, based in large part on the gaps identified here.

1.3 Structure of the document

This document is in two parts:

- a high-level summary of the state of the art in a number of areas related to traveler response to information. It attempts to summarize what is known in the area, and also to point out major gaps in current knowledge; and

- a series of reviews (annotated extended abstracts) of relevant documents. These documents provided the knowledge and data that were used in preparing the high-level summary.

The summary discussion covers:

- traveler behavior without information (Section 2);

- traveler behavior with information (Section 3);

- network impacts of ATIS (Section 4); and

- modeling ATIS network impacts (Section 5).

In the document reviews, a single review sometimes covers several documents because of their logical or organizational connections; frequently these are cases where a series of articles describes a line of research pursued over time. To facilitate locating particular document reviews, a listing is provided following the table of contents; it references each document with the number of the page where it is reviewed.

2 TRAVELER BEHAVIOR WITHOUT INFORMATION

Before beginning a review of the literature on the effect of information on traveler decision-making, it is worthwhile to briefly summarize current approaches and understanding of such decision-making in the absence of information. This is useful for a number of reasons.

Traveler behavior exhibits many features that do not depend in a significant way on whether information from external sources is available or not. Many of these features have been identified and elucidated through studies of behavior without external information. Furthermore, it is likely that many aspects of traveler behavior in the presence of information are variations on similar behavior without information. For example, if travelers are sensitive to travel time in selecting their travel path, it is likely that many aspects of their behavior when they have reliable information on travel times will be similar to their behavior when they had to estimate these times. However, the availability of more precise and reliable time estimates may lead to modified or new behaviors that were not present when only low-quality information could be had.

Understanding of the factors that travelers consider when making trip-related decisions, and of the relative importance of these, can suggest which types of information an ATIS should provide.

Many of the methods that have been developed over the decades to analyze traveler behavior in the absence of information remain applicable to the analysis of behavior with information, so it is worthwhile to briefly review these in the simpler no-information context.

Finally, in some ATIS technologies, travelers will make portions of their trips without information and other portions with information. Consider an ATIS that transmits traffic information over a short range only: a VMS or low power radio transmitter, for example. A driver might leave home having made her travel decisions without input from the ATIS, and only receive reliable real-time information in the middle of her trip. The trip thus consists of two portions: an initial segment without information, and a final segment with information. Accurate predictions of driver behavior and of the network impacts of ATIS would require reliable models of decision-making in both contexts.

In short, traveler behavior with information cannot be understood without knowing something about traveler behavior without information.

2.1 Route choice

Transportation professionals since the beginning have had to consider the question of traveler route choice behavior, since it directly affects network-level traffic flow patterns and costs. For simplicity and convenience, any analyses have assumed that travelers choose, from among a set of alternative routes under consideration, the one that offers the lowest travel time or travel cost. From introspection and observation, however, it is not difficult to conclude that this is usually only an approximation of a more complex decision-making process.

There have been many efforts over the years to obtain a more detailed understanding of how travelers decide which routes to consider and then select one to follow. Many of these have been directed towards understanding the decision mechanism that underlies travelers' route choice behavior and establishing an appropriate modeling theory and modeling form. A number of selected research articles were reviewed to highlight some of these modeling efforts and the methods they employ.

One of the basic approaches to understand drivers' route choice behavior is descriptive data analysis. Data collected in the field and from driver surveys are used to infer drivers' route choice criteria and their relative importance drivers' decision-making processes. Descriptive statistics of the data form the basis of this approach. (Huchingson, McNees et al. 1977) and (Ratcliffe 1972) used this kind of approach to find the driving habits of the drivers – routes taken, reasons for selecting these routes, and the most important factors influencing the selection. (Heathington, Worrall et al. 1971) conducted a similar study. They found that drivers were more likely to divert to avoid delays or to save travel time on the trip to work than on the trip home; they further found that drivers were more likely to divert in order to avoid delay rather than to save travel time.

Another distinct approach in the existing literature is to use different statistical techniques like principal component factor analysis, canonical correlations, multiple regressions and grouping techniques. (Wachs 1967) used principal component factor analysis to determine whether different reasons that individuals gave to explain their route choices indicated the same or different underlying values. Respondent's attitudes were examined to determine whether they were influenced by the performance characteristics of the routes. Statistical explanation of the attitudes, in terms of driver and route characteristics, was approached by three methods: canonical correlation, multiple regression and grouping techniques. The results of these analyses are presented and conclusions are drawn regarding the dependence of attitudes toward route choice upon persons and route characteristics. (Heathington, Worrall et al. 1971) also conducted a factor analysis to determine whether relationships existed between diversion frequency and other selected respondent characteristics. However, they did not find any meaningful relationship. (Pedersen 1998) used principal component factor analysis to identify the factors that influence person's route choice. Four orthogonal factors involved in selecting automobile routes were obtained: safety, interest, purpose and hindrances. A profile analysis was also performed to find if these factors were differentially rated by men and women.

Route choice can also be modeled as a continuous variable in a variety of ways. (Duffell and Kalombaris 1998) identified the main route serving various trip origins and destinations, then used regression analysis to estimate the percentage of drivers using a route other than the main route under consideration.

Disaggregate (i.e., individual-level) choice analysis methods based on random utility models have been widely applied to model drivers' decision making processes. In the context of disaggregate route choice modeling, the routes available to a traveler make up the choice alternatives, and the model predicts the probability that each of the routes in the set will be chosen. In this class of models, simple multinomial logit models are the simplest and perhaps most commonly used. However, the IIA (independence from irrelevant alternatives) property of the simple logit model restricts its applicability to general route choice analysis. This property results from the logit model assumption that path utilities include a random error term, and that the error terms of different paths are statistically independent of each other. Particularly in urban road networks, where alternative paths may overlap over significant portions of their length, the IIA property can be violated because of correlations in unobserved path attributes.

A number of modifications to the basic multinomial logit specification have been proposed to address this problem. For example, a size variable or a commonality factor may be included in the utility function to account for overlap between paths in the choice set. Another approach is the scaled paired combinatorial logit model, which scales the path utilities by a pair-wise similarity parameter. These models retain much of the simplicity and computational convenience of the basic logit model form, but overcome the unrealistic consequences the IIA property by coping with the correlation between paths.

The nested logit model, a generalization of the simple logit model, has also been used for route choice modeling. The advantage the nested logit model is that, by construction, it avoids the IIA property of the standard logit model. Estimation of nested logit models is only slightly more complex than that of simple logit models; software is readily available for this purpose.

Application of discrete choice modeling methods to route choice behavior is made complicated by the very large number of practically feasible routes between most origin and destinations, and the complex overlapping of these routes. The paper by (Ben-Akiva, Bergman et al. 1984) treats these difficulties by developing a two-stage model structure: choice set generation followed by selection from the choice set.

In the first stage, a labeling approach is used to reduce the huge number of potential routes to a much smaller number of routes, each of which reflects a criterion that might be relevant to route choice. These criteria (minimize time, minimize distance, maximize scenery along routes, etc.) are called labels. For each label, a criterion (or a generalized impedance) function is defined so that a network minimum path algorithm can be used to build trees that are minimal with respect to the criterion. Paths in these trees emphasize the corresponding label characteristics. For example, when considering the scenery label, time spent on roads with poor scenery would be weighted much more heavily (i.e., have greater impedance) than time on scenic roads. In specifying and selecting these labels, the objective is to generate a reasonable set of paths that include the actual paths chosen by the drivers. The selection of labels is made to maximize the coverage by the label set of the actually chosen paths, and the optimal values of the parameters of the impedance functions are the values that maximize this coverage. A deterministic choice set generation model is estimated for this purpose.

In the second stage, a model of choice from the set of labels is applied to predict the chosen route. A discrete choice model in the form of nested logit model is used for this stage. Path attributes specified in the utility function include generic variables like time and distance that describe the physical path, as well as dummy variables. The resulting model formulation was too complicated to be estimated using available software. Estimations were made with a series of successively less severe restrictions imposed on the general model.

In the study of individual route choice behavior, it is important to capture the heterogeneity in drivers' tastes (preferences). In general, taste variations across individuals results in differences regarding their responses to alternative attributes and their preference to various choices. Similarly, when studying the behavior of an individual over time (because of repeated surveys, for example, or when modeling a learning process), it is important to recognize potential correlations between the individual's choices. A logit model with fixed coefficients is not capable of fully accounting either for the variations in taste between individuals or the correlation between repeated choices by the same individual over time. Accurate modeling of route choice behavior requires a model that can capture differences in intrinsic preferences and subjective evaluation of alternative attributes due to both observed and unobserved heterogeneity.

The mixed multinomial logit (MML) model provides the flexibility to cope with these issues. In the MML model, an additional error term is added to the utility specification. Depending on the model, the additional error term may have a normal, uniform, log-normal or other distribution, with parameters to be estimated. The additional term captures heteroscedasticity among individuals and allows correlation over alternatives and time. However, this generality comes at a cost: choice probabilities cannot be computed analytically as they can, for example, in a logit model. Simulation techniques must be used to approximate the choice probabilities needed for model estimation and application. Recent advances in simulation-based estimation procedures make this more computationally feasible than it formerly was.

(Han, Algers et al. 2001) used an MML formulation to model route choice. Different error term distributions and model specifications were tested. The models with log-normal error distributions could not be estimated due to computational difficulties, leaving three alternative distributions – fixed, normal, and uniform. The logit model tested with fixed coefficient values differs from the standard logit model by incorporating the correlation between repeated choices by an individual. Dramatic improvement in the statistical performance of the models was found by allowing the coefficients of observed variables to vary randomly across individuals. The change in the estimated parameters caused by using the MML model was also significant. Parameter coefficients are generally larger in the MML relative to the simple logit model.

2.2 Departure time choice

Peak period congestion is one of the most persistent problems facing the transportation system. Transportation planners and transit operators have become increasingly aware of the need to spread the concentration of peak period travel. Various strategies proposed to combat the peak period problem are based on encouraging commuters to alter the time at which they travel to work. One way of assessing the potential impact of these strategies is to develop an understanding of the factors that affect commuters' departure time decisions. A significant amount of research has been done on modeling commuters' departure time choice in the absence of information.

A number of research papers on this topic have been reviewed. Again, given the amount of published research and the limited time frame available for the literature review, this cannot be considered a comprehensive survey of available material; rather, it highlights a number of interesting and representative research efforts and their conclusions.

Many research efforts apply disaggregate random utility models, of which the simple multinomial logit model is perhaps the most widely used. In the context of departure time choice modeling, discrete departure time intervals are used as the choice alternatives.

Departure time was modeled in combination with mode choice by (Hendrickson and Plank 1984): mode and departure time choices were treated as a simultaneous interactive decision. They developed a logit model that included up to twenty-eight alternatives, representing combinations of four modes (drive alone auto, shared ride, transit with walk access and transit with auto access) and seven different departure time intervals of 10 minutes each. The modal utility specification included: free flow in-vehicle travel time, the portion of total travel time due to congestion; monetary cost divided by income; walking time on the home end of a transit trip; wait time; minutes of late arrival at work and a quadratic function of that; minutes of early arrival at work and a quadratic function of that.

Much departure time research has focused on auto commuters; transit users have been neglected from consideration. One exception is a discrete choice modeling study by (Abkowitz 1981) of departure time choice. Among the objectives of this research were to extend the study of commuter departure time to include transit commuters, to include consideration of a wide range of socio-demographic characteristics, to account properly for the travel time uncertainty in departure time choices, and to improve the definition of arrival measures. Departure time choice was modeled conditional on mode choice. Departure time was represented as a discrete choice, using a logit model formulation. Each alternative represented a five-minute departure time interval, and the data input for each alternative represented an average of departure attributes for the interval. It was assumed that transit service frequency was sufficiently high during the peak period that all transit users were given a full set of choices.

Although the multinomial logit model structure is appealing to researchers because of its simple formulation, its IIA property is not always appropriate. In the context of departure time modeling, the IIA property implies that that the comparison of two departure time intervals does not need to consider whether they are adjacent or non-adjacent. In reality, two adjacent intervals are likely to be perceived similarly due to unobserved attributes common to both.

The ordered generalized extreme value (OGEV) structure generalizes the MNL structure by allowing an increased degree of sensitivity between adjacent departure time alternatives compared to between non-adjacent departure time alternatives and avoids the IIA restriction. (Steed and Bhat 2000) attempted to model departure time choice using an OGEV structure. However, the dissimilarity parameter in the OGEV model was greater than 1, implying inconsistency with utility-maximization theory. Hence, only the MNL structure was used for the analysis.

The argument in support of the treatment of departure time as a discrete choice is that travelers can only distinguish among a few prevailing traffic conditions over a specified departure period. However, discretizing departure time imposes an arbitrary structure of time intervals on the decision model. (Abu-Eisheh and Mannering 1987) develop and estimate a model that treats departure time as continuous variable and thereby avoids any *a priori* restrictions due to time discretization. Departure time is modeled as a function of the work start time, travel time, work access time and delay cushion (defined as the time difference between work start time and arrival time). Work start and work arrival times are assumed to be exogenous to the route and departure time choices. Travel time on a route is modeled as a function of route specific characteristics, commuter socio-economic characteristics and vehicle characteristics. However, since travel time on a route and the route choice are interrelated, there is a selectivity bias. The expected value method is used to correct this problem, where every route specific variable included in the travel time equation is replaced by its expected value. Delay cushion on a route is also modeled as a function of route specific characteristics, commuter socio-economic characteristics and commuter preferences for early or late arrival. The delay cushion model is also corrected for possible selectivity bias. The travel time and the delay cushion models are estimated by ordinary least squares.

Another approach to departure time modeling uses Poisson regression. The motivation for this is the belief that commuters never completely settle on a fixed departure time and route because they continually experiment with travel options and because of random effects such as weather. Within this context, a Poisson distribution is found to be a reasonable description of the number of departure time changes. Such a methodological approach is commonly referred to as Poisson regression. (Mannering 1989) and (Jou and Mahmassani 1994) used this approach to model the number of departure time changes by commuters within a month and a week respectively.

A novel approach to model driver departure time decisions is to investigate the cognitive aspects of the decision. This approach treats the departure time choice as a problem of decision-making problem under uncertainty. It criticizes the expected utility theory approach that is frequently applied to departure time modeling because expected utility theory is felt to ignore the cognitive processes underlying observed travel behavior. Depiction of travel behavior under uncertainty requires cognitive models, rather than probability theory, to capture the mental representation of uncertainty. Another finding of this kind of approach is that the decision frame, i.e. the subjective interpretation of the decision problem, critically affects decision-making. It has also been pointed out that the uncertainty of outcome is perceived as an interval of possible resultant values. Based on these findings from cognitive science, (Fujii and Kitamura 2001) propose a model of commuter departure time choice based on a cognitive task and a mental representation of uncertain travel time. By using departure time choice data, the study shows the presence of decisional phenomena, which are poorly explained by expected utility theory, but are explained well by the proposed model.

Most of the research on departure time modeling considers peak period work trips exclusively. In contrast, (Steed and Bhat 2000) modeled departure time choices for home-based recreational and shopping trips. This research examines the effect of socio-demographic characteristics, employment-related attributes, and trip characteristics on individuals' departure time choices. The departure time alternatives are represented by several temporally contiguous discrete time periods such as early morning, a.m. peak, a.m. off-peak, p.m. off-peak, p.m. peak, evening. The choice among these alternatives is modeled using a discrete choice model. Two alternative discrete choice structures were explored. The first is the multinomial logit (MNL) structure and the second is an ordered generalized extreme value (OGEV) structure.

2.3 Mode choice

The literature on mode choice modeling is vast, and no attempt was made to review or summarize it. The following paragraphs simply note some modeling approaches commonly applied.

As travel modes are by their nature discrete alternatives, discrete choice models suggest themselves as a natural modeling approach. In this approach, all the modes available to a traveler constitute the choice set. Simple logit models are often applied to compute the probability of choosing each mode. The utility to a traveler for a particular mode can be a function of travel time (in-vehicle and out-of-vehicle) on that mode, out-of-pocket costs on that mode, perceived costs on the mode, socio-economic and demographic characteristics of traveler, workplace dummy and lots of other dummy and continuous variables. Many of these variables can be specified either generically or as specific to one alternative.

As has been mentioned above, the standard logit model has the independence from irrelevant alternatives (IIA) property. This means that for a specific individual the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utility of any other alternative. This can be unrealistic in mode choice modeling, because some modes in the choice set may have similar unobserved attributes and so have correlated utilities. An individual choosing between auto, commuter rail and express bus, for example, is likely to have somewhat similar (positive and/or negative) feelings about bus and rail, so treating them as completely independent vis-à-vis the auto could lead to unrealistic choice predictions.

The simplest generalization of the logit model that avoids this problem is the nested logit model; properly specified, it does not suffer from the IIA property. In this modeling approach, alternative modes that are likely to have unobserved common attributes should be put in a single nest and the resulting model should be used. The model incorporates a higher-level choice between nests, and a lower-level choice among the alternatives in a nest. In the previous example, it would be reasonable to group the commuter rail and express bus in a single "commuter transit" nest. The high-level choice would be between auto and commuter transit, with a lower-level choice between bus and rail in the transit nest.

3 TRAVELER BEHAVIOR WITH INFORMATION

This section considers the question of traveler behavior in the presence of real-time travel information.

This general question actually involves a number of closely inter-related sub-questions:

- which kinds of travelers would use real-time travel information if it were available? What kinds of trips would they want to use it for?
- how would they respond to the information once they received it? How would directly affect decisions about a trip being contemplated or made? How would it affect the context in which trips are made?
- what specific types of information would these travelers want to access?
- how much would they be willing to pay to receive the information?
- what would be their assessment of the benefits they received from accessing the real-time travel information and responding to it?

- how would this assessment of their experience affect the answers to all these questions the next time they have the opportunity to use it?

Because of their deep interdependence, all these questions should ideally, perhaps, be addressed and answered simultaneously. However, it is necessary to begin somewhere. Therefore, this section starts with a review of some of the literature that analyzes and characterizes the potential users of ATIS. From this, it turns to examine the various kinds of user response to travel information that have been studied. It then looks at users' preferences and willingness to pay for different types of information. There follows a discussion of the dynamic effects that can occur when day-to-day learning behavior is considered. Finally, a number of specific topics in traveler response data collection, analysis and modeling are discussed.

3.1 Who are the potential users of real-time travel information?

Understanding who are the potential users of advanced travel information services is essential both for designing and marketing those services and for predicting the users' responses to them. It is intuitively clear that ATIS can serve a variety of different kinds of users, and that these different kinds of users may react to ATIS messages in substantially different ways. The better these differences are understood, the better user needs can be met and user response can be predicted.

Studies of travel behavior are increasingly drawing on ideas and methods of market research. These methods typically attempt to identify subgroups ("segments") of the total market having the property that individuals within a subgroup share many similarities with respect to variables of interest in a study (e.g., travel behavior, socio-economic characteristics), and individuals in different subgroups differ significantly along these dimensions. Each homogeneous market segment can be more efficiently studied than can the mixed population as a whole.

A straightforward way of implementing these ideas is to identify segments on the basis of the exhibited behavior of interest (e.g., ATIS users), and to correlate membership in the segment with other measurable characteristics (e.g., socio-economic characteristics). Although useful, this approach has the disadvantage of being able to identify only relatively simple correlations, and perhaps also of reflecting the analyst's a priori beliefs and preventing a more exhaustive exploitation of the data.

More sophisticated market research methods such as cluster analysis can statistically identify population subgroups whose members share high degrees of similarity across many dimensions. While outputs of statistical procedures always need to be interpreted with insight and caution, clustering methods are often capable of identifying previously unknown significant population segments that might not have otherwise been recognized in the data.

Factor analysis is another method of identifying structure in a data set consisting of multiple observations, each one involving multiple variables of interest. Factor analysis identifies sets of linear combinations of the variables that distinguish as much as possible among the observations. Given a particular linear combination of variables (a *factor*), an observation's *score* with respect to the factor is the numerical value of the linear combination evaluated using the particular values of the observation's variables. Factor analysis identifies factors such that (i) the distribution of scores with respect to each one has maximum variance (i.e., the factors have maximum discriminatory power), and (ii) different factors are orthogonal to (i.e., uncorrelated with) each other. When a factor's linear combination includes some variables with very high coefficients and others with very low coefficients, its interpretation may be relatively easy. Factors involving more general linear combinations with arbitrary coefficients on the variable may be more difficult to interpret. In such cases, identified factors may subsequently be "rotated" to facilitate their interpretation in terms of specific variables or sets of variables, and this rotation may introduce correlations between them.

The combination of factor and cluster analysis is a particularly powerful means of identifying market segments, and has come to be a standard method in market research. Factor analysis is first applied to a data set of survey results to identify a set of factors that efficiently and parsimoniously distinguishes the observations. Each observation's scores with respect to the different factors are computed, and then cluster analysis is applied to identify subgroups of observations having similar factor scores. It remains for the analyst to impose a meaningful interpretation of the subgroups so obtained.

(Proussaloglou, Haskell et al. 2001) describe an application of combined factor and cluster analysis to identify transit user market segments in the San Diego metropolitan area. They then develop (fairly conventional) transit mode choice models for each distinct market segment.

Turning to analyses of the potential market for ATIS services, most surveys of potential ATIS users have carried out simple correlations or other descriptive analyses of stated use propensity with socio-economic or characteristics. Work pursued over a number of years by a group at the University of Washington (Barfield, Haselkorn et al. 1989; Haselkorn, Spyridakis et al. 1989; Haselkorn, Barfield et al. 1990; Wenger, Spyridakis et al. 1990; Spyridakis, Barfield et al. 1991; Conquest, Spyridakis et al. 1993) is among the first examples of the application of cluster analysis techniques to investigate the characteristics of potential ATIS users. Based on an mail-in driver survey and follow up personal interviews, the researchers were interested in the respondents' use of traffic information (commercial radio and TV traffic reports, HAR, VMS) and response to it, and in the influences that affect these responses. Cluster analysis of the survey results was intended to identify subgroups that differ significantly in their use of traffic information. The four groups identified by the cluster analysis were (in decreasing order of frequency in the sample): departure time and route changers; non-changers; route changers; and pre-trip changers. (Although mode change behavior in response to travel information was also investigated, the number of respondents who reacted to travel information by changing mode was not significant.) Descriptive statistical analysis was then used to further characterize each of the identified market segments in terms of its use of and attitudes towards different information sources; its priorities with respect to different information features; its tripmaking and activity constraints; and its demographics.

(Mehndiratta, Kemp et al. 1999b) (see also (Mehndiratta, Kemp et al. 2000; Mehndiratta, Peirce et al. 2000)) illustrate the application of combined factor and cluster analysis techniques, as described above, to delineate distinct segments of ATIS users. A detailed collection of data on travel behavior including use of travel information was conducted as part of the ongoing Puget Sound Regional Council's travel diary panel survey. The survey included conventional demographic and socio-economic information as well as responses to attitudinal questions. From this data, individuals with a high propensity to use travel information were identified. An initial attempt to correlate membership in this group with socio-economic characteristics, based on stereotypes of expected users types (e.g., road warriors, commuting mothers) proved only partially successful. Accordingly, a factor analysis of the entire survey population’s attitudinal question responses was performed, and a cluster analysis using the factor scores was carried out to identify distinct segments. Although the segments were defined uniquely in terms of their attitudes, subsequent analysis showed that the segments also differed with respect to their travel behavior, demographic profile, and propensity to use ATIS. The incidence in each segment of individuals likely to use ATIS was then determined.

Eight distinct market segments were identified through the combined factor/cluster analysis. The segments with higher-than-average incidence of ATIS users were termed:

- control seekers: people who travel a lot, are comfortable with technology, like to plan ahead but are not set in their ways;

- web heads: people who are interested in cutting-edge technology and traffic information, although they are less interested in portable electronics.

- rigid routines: people who usually follow the same routine but listen to traffic information and will make small adjustments to their trips;

- value-added service buyers: people uncomfortable with maps and computers who appreciate things that facilitate their daily lives;

- wired with children: people with high incomes, long commutes and children, for whom convenience is important.

Subsequent application of this approach to a wider sample of people who had used ATIS during the various MMDI programs revealed an additional potentially important segment:

- mellow techies: people with little interest in traffic conditions or trip planning, and little concern about being late, but who have high levels of internet and computer use.

It is clear that application of techniques such as these can provide considerable insight into the structure of the market for ATIS services, and allow much more focused investigation of the characteristics, system preferences and behavioral responses of potential ATIS users.

3.2 Traveler response to real-time information

(Polydoropoulou and Ben-Akiva 1999) have described a number of successive stages that travelers typically go through before they become regular ATIS users. These are:

- awareness, where the traveler begins to have basic information about the availability and attributes of a travel information system;

- consideration set formation, where the traveler generally begins to think of ATIS as a possible option to consider before making trips;

- choice set formation, where ATIS is definitely included as an option to assess in response to a specific identified travel need;

- trial use, where the traveler decides to try ATIS to gain more familiarity with its characteristics and potential benefits and costs;

- repeat use, where ATIS is assimilated into a traveler's continued or habitual travel behavior, although further experience may cause the continued use to be reconsidered.

At the point where repeat usage becomes established, it becomes possible to speak of a systematic traveler response to real-time information. These responses are divided here into two general categories: those involving the tripmaking context, and those involving tripmaking itself. The sections below discuss these responses, drawing on the literature review to indicate the extent of current qualitative and quantitative knowledge about the responses.

Responses to ATIS involving the tripmaking context include behavior that affects the way that trips are scheduled or integrated into daily activities. These include adjustments to residential and/or employment location decisions; adjustments to daily activity schedules; changes in habitual tripmaking behavior; effects on non-travel activities; and trip-related stress or anxiety relief.

Responses to ATIS involving tripmaking itself cover a wide range of trip-related decisions: the decision to travel or not; the choice of destination or destinations (trip chaining); choice of departure time, mode and route; the re-routing decision in response to an incident; driving behavior; and the choice of parking location.

These various possible responses are discussed individually below, despite the fact that in many cases the responses are inter-related. The discussion also examines a number of specific examples of traveler response that merit separate consideration; these include ATIS impacts on shopping trips, transit information systems, variable message signs, and driver compliance with prescriptive information.

It will be seen that, in most cases, very little quantitative information is available. The available information tends to be highly specific to particular situations; very few quantitative conclusions of a generally applicable nature can yet be drawn regarding user responses to ATIS. This is not entirely surprising: significantly research into and deployment of ATIS has only been taking place for the past decade or so. Highways and transit systems were in use for many decades before generally reliable data and models on traveler response to them began to be developed. The pace of research and investigation is faster now, and the methods of data collection and analysis more efficient and sophisticated. Still, the current state of knowledge provides at best general qualitative conclusions regarding traveler response to ATIS. More deployments, more experience with deployed systems, and more research and analysis will be required to move ahead.

3.2.1 TRIP CONTEXT RESPONSES TO ATIS

3.2.1.1 Reduce stress and anxiety

Many surveys have found that tripmakers appreciate having travel information available even if they do not or cannot modify their tripmaking behavior in any way because of it. Some analysts see this reaction as similar to peoples' appreciation of weather forecasts. Respondents typically claim that the information reduces the level of anxiety or stress associated with not knowing what travel conditions are going to be. (Khattak, Schofer et al. 1995) and (Khattak, Yim et al. 1999), for example, discuss survey results where users mention this reaction.

(Lee 2000) has attempted to make the notion of travel stress relief more precise by arguing that the value of time spent in travel includes at least two distinct components: the opportunity cost of the activities foregone by traveling, and the disutility of the travel experience itself. This disutility is likely to be higher when a lack of information about travel conditions ahead causes one to be anxious or under stress; conversely, receiving travel information may make one more "serene" during a trip. The value of time spent traveling is likely to be higher in the former case than in the latter, and the benefit of the stress-relieving impacts of ATIS can be estimated as a function of the difference in value of time and the total time spent traveling.

3.2.1.2 Affect non-travel activities at the trip endpoints

Travel information may enable tripmakers to beneficially adjust the activities that they undertake at the departure or arrival ends of a trip. A person stuck in traffic may be able to call ahead with an accurate arrival time estimate and, before arriving, re-arrange her schedule at the destination to minimize the impacts of the delay on other activities. A person who wants to complete a task at one location but also needs to arrive at another location on time may be able to make use of accurate travel time information to determine if there is sufficient time to complete the task before departing. In the absence of such information, the person may abandon the task even if there was enough time to complete it; or complete it, and arrive late at the next location.

A Mitretek study ((Shah, Toppen et al. 2001; Wunderlich, Hardy et al. 2001); see also (Shah, Wunderlich et al. 2001)) provides evidence from simulated yoked driver experiments involving the Washington DC and Minneapolis/St. Paul metropolitan areas that pre-trip ATIS can significantly reduce the early and late schedule delays, and reduce the number of late arrivals. These studies compared the travel time and arrival time reliability of pairs of simulated drivers with identical origin, destination and desired arrival time at the destination. One driver was assumed to have access to pre-trip ATIS information on link travel times, and the other not. (The link travel time information was empirical data, compiled by polling an on-line traffic information service for conditions at five-minute intervals over a large number of days.) Drivers without access to information were assumed to base their path and departure time choices on average link conditions experienced over time, while those with access were assumed to utilize the “real-time” (but non-predictive) link times to make these decisions. In each case, the consequences of the decisions, in terms of travel and arrival time, were determined by reference to the compiled data on actual link times. (Compiled values were slightly perturbed to account for the variability in the time estimates.)

The study found that pre-trip information had only a minor effect on the average travel times experienced by its users. However, ATIS users reduced their number of late arrivals by 62%, and the total late schedule delay by 72%. (These benefits varied significantly by time of day.) The conclusion is that pre-trip ATIS is likely to impact travel time reliability much more than travel time itself. The study also suggests that, in the travel contexts considered, pre-trip ATIS is more likely to produce departure time changes than path choice changes.

3.2.1.3 Adjust daily activity schedule

People schedule the activities that they need to accomplish in a day based in part on the time taken by each activity and the time required to travel between activities in different locations. Because of uncertainty about travel times, people tend to incorporate “slack” in their scheduling decisions to reduce the risk of schedule disruptions due to worse-than-expected travel conditions.

Reliable information on travel times and traffic conditions will allow people to eliminate some of this slack. The time freed up in this way could be used in a wide variety of ways. At one extreme, it could be used to sleep or relax more; at the other, it could lead to a significantly different organization of the day's activities including new activities and shifts in the order of activities. In terms of tripmaking, the additional time could lead to new trips, to trips made at different times, or to trip chaining.

Although these kinds of behavioral adjustment are entirely plausible, there is as yet very little evidence that they have occurred among users of currently-deployed ATIS.

3.2.1.4 Adjust habitual tripmaking behavior

There is considerable evidence that tripmakers rely to a large extent on habit when making their travel decisions. Over time, they establish a set of default behaviors that influence their tripmaking behavior on particular trips. These default or habitual behaviors do not necessarily dominate the decision-making process; rather, their effect is to increase the likelihood that, in any particular decision context, the default choice will be made. (Aarts, Verplanken et al. 1997) provide an analysis of bicycle use by students that supports this view.

(Uchida, Iida et al. 1994) surveyed commuters in a three-route corridor in Osaka, Japan following the installation of a VMS network that provided predicted travel time information. They identified two types of response to the information:

- tactical response, meaning the immediate decision to divert or not based on reported travel times for the three routes; and

- strategic response, the change over time in drivers' selection of their habitual route.

The VMS was found to significantly affect both types of response. However, decision inertia was also found to be important in both. In the case of the tactical response, drivers showed a reluctance to switch away from their habitual route, other things being equal. In the case of the strategic response, drivers were reluctant to change their habitual route, even when the VMS repeatedly showed it to be an inferior alternative.

(van Berkum and van der Mede 1998) present a sophisticated modeling and analysis framework that accounts for the effects of ATIS in immediate travel decision-making and longer-term habit formation and change. The article presents empirical results that support their framework and highlight the importance of habit in tripmaking behavior. Similar results are presented, in another problem situation, in (van Berkum and van der Mede 1999).

One potentially important factor not considered in these studies is the possible effect of ATIS-produced changes in the daily activity pattern on the formation of travel habits. If, as was discussed in the preceding section, accurate travel condition information from an ATIS leads to a reorganization of a persons' daily activity pattern, it is probable that habitual travel behavior will also change as a result.

3.2.1.5 Adjust residence and/or employment location

The variety of changes brought about by ATIS in the tripmaking context could lead people to reconsider their decisions regarding residential and/or employment location. As one example, if more predictable travel times became available from an ATIS, households could move farther away from job locations while still maintaining the same average commute time. Again, rearrangements in daily activity schedules brought about by ATIS could allow more time for outdoor activities, and incite households to take advantage of this by moving. Through these kinds of effect, ATIS could ultimately have an impact on urban form and structure. (Boyce 1988), in an early paper, evoked this possibility. (Hamerslag and van Berkum 1991) presented a simple network model that exhibits such location decision effects. However, it is likely that ATIS deployment on a much larger scale than today's will be required before such effects become noticeable or significant.

3.2.2 TRIPMAKING RESPONSES TO ATIS

3.2.2.1 Decision to travel or not

Relatively little information is available regarding the effects of ATIS on the decision to travel or not; however, it is not inconceivable that information about sufficiently bad travel conditions could induce tripmakers to cancel their intended trips, particularly discretionary trips.

(Khattak, Yim et al. 1999) cite evidence for this effect from CATI and mail questionnaire surveys carried out as part of the San Francisco-area TravInfo project. The surveys covered automobile and transit travelers and commute and non-commute (e.g., shopping or personal) home-based trips. The surveys asked respondents about the effects of pre-trip travel information (available from television, radio or telephone sources) on their tripmaking decisions. Analysis of the survey results revealed a number of general aspects of traveler response to the available information sources, some of which are discussed in sections below.

One of the findings was that non-commuters would occasionally decide to cancel their (presumably discretionary) trips because of unfavorable travel conditions reported by the various information sources, and particularly by radio. It is widely agreed that the demand for non-commuting trips is relatively elastic with respect to travel times and costs – in other words, an increase in travel times or costs leads to a reduction in tripmaking. In view of this, it is not surprising that information about bad travel conditions would lead, at the individual level, to non-commute trips being canceled. However, this is the only empirical evidence that was encountered in the literature review for such an effect.

3.2.2.2 Choice of destination or destinations

Similarly, relatively little information is available in the literature regarding the effects of ATIS on destination choice, or on the decision to visit several destinations and accomplish several purposes in one trip through trip chaining. Trips offering a choice of destination alternatives are likely to be for shopping or personal purposes, rather than for commuting. The opportunities to group multiple purposes and destinations into a trip chain are more varied and difficult to characterize and analyze.

The effects of ATIS on shopping trip destination choice was investigated in a set of internet-based stated preference surveys by (Kraan, Mahmassani et al. 2000) and (Mahmassani, Huynh et al. 2001). In the survey, respondents were asked to make a (simulated) shopping trip from a central location in Austin, Texas to a major suburban mall. Different pre-trip and en route messages about travel conditions were provided in the course of the decision-making process. Following notification of a change in traffic conditions while en route, the respondent was given the options of continuing on the same route; continuing to the same mall but via a different route; or switching to a different shopping mall entirely. Appropriate information was provided in each case. A sequential decision framework was developed to capture the conditional nature of the choices. It was found that the decision to switch route or destination was not influenced by age, gender, education and income. Respondents who were less familiar with the Austin area were more likely to switch destination, but not route. Those who visit the same mall on a frequent basis were less likely to switch destination and route. In general, switching response was greatest when information on traffic delays (as opposed to other kinds of traffic data) was presented.

Again, these are the only references located during the literature search on the topic of destination choice and trip chaining impacts of ATIS. Indeed, these questions are not well covered in the broader transportation literature; data on trip chaining, in particular, is difficult to collect and analyze.

3.2.2.3 Departure time choice

Departure time and route choice are often considered together in discussions of travel behavior. Many surveys of pre-trip user behavior collect data on both types of decision. They are considered separately in this discussion of ATIS because route choice can potentially be influenced by both pre-trip and en route information, whereas departure time choice is by its nature a pre-trip decision only.

There are a number of indicative data elements regarding the influence of ATIS on departure time choice but, again, the available data is not complete enough to draw broadly general conclusions or to develop widely applicable models.

An early study of commuting behavior (Mahmassani and Chang 1985; Mahmassani and Chang 1986) gave some indication of the slack that commuters feel they need to build into their departure times. Around 40% of survey respondents stated that they schedule their commute trip to arrive at work at least 15 minutes before the official start time; furthermore, the early schedule delay was found to increase with increasing distance from work. This suggests that travel time variability influences the departure time decision, and that commuters leave their homes early in order to reduce the risk of late arrival from longer-than-expected travel times.

(Barfield, Haselkorn et al. 1989) (Haselkorn, Barfield et al. 1990) (Mannering, Kim et al. 1994) discuss results of surveys of Seattle-area commuters who receive travel information from radio, television and telephone services. Of the commuters surveyed, 40% indicated that they had some flexibility in scheduling and selecting the route for their morning commute trip; 23% indicated no flexibility. However, 64% responded that they rarely changed their departure time because of pre-trip information.

(Khattak, Schofer et al. 1991) and (Khattak, Yim et al. 1999) report that the perceived accuracy of pre-trip reports is important in determining whether commuters take account of it in their decision-making. The importance of perceived pre-trip accuracy was also reported by (Polydoropoulou and Ben-Akiva 1999) based on analyses of San Francisco commuter surveys.

(Srinivasan and Mahmassani 2001) investigated using travel choice simulators the mechanisms by which drivers arrive at a departure time decision based on ATIS messages. They hypothesized that a driver undertakes a sequence of decisions to arrive at an adjustment to her habitual departure time. First, the driver decides whether or not to adjust the habitual departure time. Conditional on the decision to adjust, departure time alternatives are evaluated sequentially in about five minute increments. The directionality of adjustment (i.e., towards earlier or later departure) is governed largely by the direction of schedule delay experienced on the preceding day, with an earlier switch following prior lateness and vice versa. The results illustrate that the observed departure time adjustment behavior is influenced by dynamic transportation system attributes encountered such as trip time variability in the network, trip-makers' short and longer term experiences, and the nature, type and quality of real-time information supplied by the ATIS.

3.2.2.4 Mode choice

Relatively little detailed information is available about the mode choice impacts of ATIS, although there is some evidence for this effect.

As reported in (Yim and Miller 2000), less than 1% of the early callers to San Francisco's Travinfo service asked to be rerouted to the transit menu after learning about bad traffic conditions from the traffic menu. However, as experience with the system increased over the duration of the Travinfo field test deployment, it was found that up to 5% of the callers asked to be rerouted to the transit menu, a significant increase. Of those who accessed transit information, 90% of them chose transit for their travel mode. (Of course, a large fraction of the callers probably consisted of habitual transit users; it cannot be concluded that the information that they received caused them to choose transit.)

(Polydoropoulou and Ben-Akiva 1999) (see also Khattak, Polydoropoulou et al. 1996) discuss an analysis of San Francisco data that showed that prescriptive recommendations to take public transport have a detectible effect on mode choice, particularly in situations of unexpected delay.

3.2.2.5 Route choice

Many surveys and travel choice simulator studies have demonstrated the ability of ATIS to influence route choice. (Khattak, Yim et al. 1999), for example, presented survey results in which over 50% of respondents reported that they had made travel route or departure time changes in response to pre-trip information received by radio, television or telephone. (Owens 1980) describes an early travel choice simulator study that demonstrated drivers' willingness to divert in response to highway advisory radio (HAR) messages about incidents. Some researchers have estimated sophisticated econometric models of route choice or route switching probabilities in response to ATIS, for example (Uchida, Iida et al. 1994) and (Polydoropoulou and Ben-Akiva 1999).

However, as stated above, from these various surveys and modeling efforts it is difficult to extract generally applicable quantitative conclusions regarding traveler response to information. The state of knowledge does not yet allow the development of a general model capable of predicting that, on a given network, X% of drivers will divert to route Y if they receive message Z while driving. Unfortunately, sufficient experience with and data about these systems is still lacking. Accordingly, this section will focus on qualitative conclusions that have been obtained from the various analysis efforts that were alluded to above.

Based on analysis of driver route choice responses to both VMS and radio information, (Emmerink, Nijkamp et al. 1996) have suggested that some people have a natural propensity to use traffic information of any kind and from any source. (See the discussion in Section 3.1

above.) Nonetheless, there is considerable evidence that the nature of the guidance information, and the conditions experienced prior to its dissemination, can strongly affect driver route choice response to it.

(Khattak, Schofer et al. 1995) and others have found that drivers tend to prefer messages that are descriptive (information about traffic conditions) rather than prescriptive (route recommendations). They found in particular that drivers are most receptive to near term predictions of traffic conditions on congested routes with rapidly changing conditions.

However, drivers' perception of the accuracy and reliability of the messages is a key determinant of their response. (Kantowitz, Hanowski et al. 1997a; Kantowitz, Hanowski et al. 1997b) have found that there exists an accuracy "threshold", beneath which drivers will simply ignore ATIS messages. Factors that increase drivers' confidence in the accuracy of the messages tend to increase the likelihood that the drivers will react to them. In the context of route choice, such factors include a driver's own observation of congestion prior (and particularly just prior) to receiving the message, and favorable experiences with the ATIS in prior uses.

Prescriptive messages do generally have an effect on route choice, as shown in many travel choice simulator studies and surveys of driver behavior. Combining a prescriptive recommendation to change routes with descriptive information justifying the recommendation has been found in travel choice simulator experiments to result in the highest route switching compliance rates. More generally, (Polydoropoulou and Ben-Akiva 1999) found that, in en route switching situations, the switching rate increased with the elaborateness (level of detail, care in justification) of the guidance messages.

(Owens 1980) found that drivers who received prescriptive information about incident diversion routes were generally more successful in avoiding the incident than those who received descriptive messages only. The success of the latter drivers depended strongly on their knowledge of the network around the incident. However, he found that the travel costs incurred by the two sets of drivers in diverting were not notably different.

(Llaneras and Lerner 2000) also investigated the ability of drivers to translate guidance messages into effective route choices. They considered "simple" and "enhanced" in-vehicle ATIS capabilities; the latter provided basic descriptive and qualitative information on incidents and congestion, while the latter provided the simple information as well as details about incidents, alternate routes, and real-time congestion conditions as well. Overall, drivers were able to use both types of system to divert around incidents. However, he also found that drivers sometimes made incorrect route choices with both types of system. The prevalence of these errors was significantly higher with the basic system; furthermore, the mistakes made with that system were generally more costly (in terms of excess delays) than those made with the enhanced system.

A number of generally idiosyncratic factors also condition a driver's route choice response to ATIS messages. A freeway bias has been observed in several studies ((Hato, Taniguchi et al. 1995; Kitamura, Jovanis et al. 1999)). Because of this bias, drivers receiving messages that suggest diverting from a non-freeway to a freeway facility are considerably more likely to switch than drivers who receive the opposite message, other things being equal.

The influence of habit or inertia on route choice response has also been noted in a number of studies ((Uchida, Iida et al. 1994) (Hato, Taniguchi et al. 1995) (Srinivasan and Mahmassani 2000b)). Drivers tend over time to establish a preferred route for particular trips. Guidance messages that suggest switching from the preferred route to another are less likely to be accepted than messages that suggest the opposite. Of course, habit does not always over-rule information received from guidance messages. A sufficiently strong message, corroborated by the driver's observations and confidence in the ATIS, will be considered. Over time, the effect of ATIS may be not only to affect particular route choice and switching decisions, but in fact to change the habitual route choices themselves.

3.2.2.6 Incident diversion response

A special case of the route choice decision occurs when a driver becomes aware, during the trip, of an incident affecting traffic conditions on the path currently being followed. Incident-related and other non-recurrent congestion is a major contributor to total congestion delays on highway networks; for example, it has been estimated that roughly half of all delays on freeways in the U.S. are due to non-recurrent causes. Driver response to an incident situation determines in part the severity of its consequences. It is expected that ATIS can be of considerable help in incident situations by providing drivers with timely information about the location and characteristics of the incident and by suggesting routing alternatives in what are, by their very nature, unexpected and unfamiliar circumstances.

The two key aspects of driver incident response are: whether the driver diverts at all; and, if the driver diverted and avoided the incident, whether she returns to the original route or continues on the alternate route. In the former case, the route switch represents a temporary detour around the cause of delay; in the later, the route switch entails choosing a completely new route to follow to the destination. The choice considerations at work in these two situations may well be different.

A number of studies have examined drivers' route diversion behavior in the presence of non-recurring congestion, applying a variety of methodologies for this purpose. This is actually one of the better-studied aspects of traveler response to ATIS, perhaps because of the natural interest in applying ATIS to alleviate incident conditions.

(Khattak, Koppelman et al. 1993) investigated factors that influence auto commuters' en-route diversion propensity. Data on propensity to divert and related factors were collected through a stated preference (SP) questionnaire survey. The effects on drivers' willingness to divert of incidents and recurring congestion, real-time traffic information, driver and roadway characteristics and situational factors were investigated using conjoint measurement.

Disaggregate discrete choice models are a natural approach for investigating drivers' diversion and return choices. Multinomial logit models (MNL) and nested logit models (that remove the undesirable IIA property of MNL) are logical model forms. (Khattak, Schofer et al. 1993) examined diversion and return choices using these two forms. The model structure represents these choices as interrelated to take account of the likelihood that drivers' diversion choices will depend, in part, on their expectation that they will or will not return to the original route. That is, the driver chooses among three alternatives: no diversion (ND), diversion and no return (DNR), and diversion and return (DR). The authors used a joint multinomial logit model of the choice among these three alternatives and a nested logit model in which the return choice is nested within the diversion decision. Both these models were estimated with equivalent systematic utility function specifications; they yielded very similar coefficient values (i.e. identical behavioral interpretation). Commuters' diversion and return behavior varied with their personal characteristics and with the characteristics of the trip they were making at the time when the choice arose. Individuals making longer trips, facing longer delays and facing less expected congestion on alternate routes were more likely to divert. Commuters who made longer trips were significantly more likely to return after diversion.

(Abdel-Aty 1998) also considered alternative logit model forms to model the three diversion options (ND, DNR and DR), in an investigation of the preferred modeling structure for the incident-related routing decision. In addition to the joint multinomial logit, two nesting structures were tested. In one of them, the DR and DNR choices were modeled under a "diversion" nest, reasoning that these choices are conditional on diverting because diversion has occurred. The other specification places the ND and the DR choices under a "maintain route" nest, reasoning that the choices are conditional on staying on the same route because the majority of the route will be the same. It was concluded that the nested logit model with the ND and DR choices grouped in a nest provided the best structural fit for the observed distribution of the routing decision in case of an incident. The superiority in this application of a nested logit structure over the simple MNL form was also established.

Use of ordered categorical response data is very common in these kinds of modeling, where the bulk of the data is obtained through stated preference questionnaires. Use of multinomial logit or probit models or linear regression may lead to biases in estimation using this kind of data. (Khattak, Koppelman et al. 1993) estimated multivariate models of diversion propensity to explore the effects of several variables simultaneously. The multivariate model used was an ordered probit model with diversion propensity as a function of delay characteristics, reported trip and route attributes and socio-economic characteristics of the respondent drivers. The ordered probit model was selected for estimation because of its ability to analyze ordered categorical response data.

Another method to investigate drivers' route diversion behavior is to analyze reported and stated data about route diversion obtained through surveys. (Khattak, Kanafani et al. 1994) analyzed a survey of commuting behavior in the San Francisco Bay Area in 1993. The questionnaire was designed to use reported diversion behavior (a measure of the true behavior) as the basis of a sequence of stated preference questions about the propensity to divert with a future in-vehicle ATIS device. This methodology increases the validity of the stated preference technique by relating the response to ATIS technology to a specific behavior that was actually practiced by the respondent. The objective of the stated preference question was to determine how incremental amounts of information provided by an ATIS device would influence the propensity to divert. It appeared that respondents overstated their propensity to divert when compared with reported behavior. Around 22% of the respondents stated that they would divert even though they reported not having diverted. On the other hand, only 5% of the people stated that they would not divert even though they actually diverted when they faced the unexpected delay. To explore the correlation between reported behavior and stated preference, a linear regression model relating the answers to each question was developed.

3.2.2.7 Driving behavior

Traveler information can be used not only to improve trip-related decision-making, but also to influence driving behavior during the trip.

For example, messages might warn drivers before they arrive at hazardous road conditions (road work, accidents, bad weather) so that they drive more cautiously. (Ng and Mannering 2000) report on vehicle simulator experiments to determine the effectiveness of such advisory messages. They developed a very realistic simulation of an actual mountain road in Washington State, and included the ability to generate fog and place snowplows in the simulation. They considered the effect on driving speed of VMS messages, in-vehicle messages and both; messages warned about the presence ahead of fog, road curves, and snowplows. They found that over short distances, the messages did cause drivers to reduce their speeds; however, over longer distances there was no noticeable speed effect. This suggests that after slowing down in response to the message, drivers drove faster in order to compensate for the delay.

(Smulders 1990) describes a subtle application of this idea. He found that merely *suggesting* appropriate freeway speeds to drivers by VMS – but with no obligation on the part of the drivers to comply – had a small but noticeable effect on average travel speeds but significantly reduced the variability in these speeds across drivers on the facility. This reduction in speed variability considerably delayed the onset of the breakdown of traffic conditions at maximum flow levels, and actually increased the capacity of the freeways where the method was applied. Speed advisory VMS are now deployed on a number of freeways in the Netherlands.

3.2.2.8 Parking search and choice

Parking guidance and information (PGI) systems inform drivers about the availability of parking at various locations or recommend facilities for use.

In general, a PGI system consists of four components:

- a counting mechanism at parking facilities to track vehicle entries and exits and thus determine facility occupancy and available spaces at a given time;
- a control center that processes data on facility occupancies and generates messages about parking availability or recommendations. Messages may also include information about other attributes of parking facilities (prices, location, etc.);
- a communications network that transmits occupancy data from facilities to the control center and disseminates messages from the control center to users;
- information access technologies by which users obtain the messages generated by the control center.

The information access technology generally consists of a system of variable message signs, arranged so that traffic traveling towards the city center receives progressively more detailed data with each VMS encountered. The messages may be based either on current occupancies or on the occupancies predicted to hold at the time a vehicle passing a VMS actually arrives at the parking facility. Occupancy data may be quantitative (actual spaces available) or qualitative ("ample space", "nearly full", "full").

A number of such systems are in use in cities around the world. In some instances, both the parking facilities and the PGI system are operated by the municipal government, but this is not a requirement. In England, for example, there are arrangements where privately-operated parking facilities provide data to a PGI system run by the local government. The hardware needed to implement the system components is commercially available.

(Allen 1993) provides a useful summary of the benefits of PGI systems. These include:

- benefits to drivers by being able to proceed directly to a parking facility with available spaces, without having to spend time searching and waiting;
- benefits to traffic and environmental conditions from the elimination of parking search traffic which, according to some estimates, can be 30% or more of all traffic on roads in city centers;

- benefits from more efficient utilization of available facilities: higher occupancy levels and increased parking revenues;

- benefits from information availability about facility usage, making possible better management of the parking system (e.g., fraud monitoring, pricing policy analyses).

Many of the issues that arise in modeling general driver response to traffic-related information also occur in modeling response to parking-related information. It appears from a review of the literature, however, that parking choice and PGI systems have been less intensively investigated to date than route choice and ATIS.

Among the articles reviewed, (Teng, Falcocchio et al. 2001) surveyed parking facility users in New York City to determine the types of information they considered most useful in a PGI system, and to investigate relationships between user or trip characteristics and the ranking of information types. For a parking information web site, the information of greatest interest included fee structure, hours of operation, location, the predicted probability of having a space available at the time of arrival, and traffic conditions in the vicinity of the facility. For roadside displays, the information of greatest interest included hours of operation, number of available spaces, location and fee structure. These preferences were observed to vary by gender, trip purpose, and familiarity with parking options and conditions, among other factors. Internet-based information and in-vehicle devices were preferred to a kiosk for obtaining pre-trip information, while VMS were preferred to in-vehicle devices for obtaining en route information.

(Allen 1993) conducted stated preference surveys in an outer borough of London to investigate the effects on parking facility choice of VMS message, parking price and walk time to the destination. The survey concentrated on weekday shopping trips and distinguished three different user groups. The considered messages provided qualitative information on the occupancy of different nearby parking facilities. Two message dictionaries were considered in the SP experiments, differing most notably in that one explicitly identified facilities as "nearly full" while the other displayed a blank message for such facilities. The authors present multinomial logit model estimation results. Within the range of prices and walk times considered, the displayed message had a determining effect on parking facility choice, while price and time had secondary effects. It is not clear, however, if these conclusions would hold over a larger range of prices and times. It should also be noted that this work did not consider the practical problem of how to indicate parking facility locations to unfamiliar drivers using a VMS with very constrained message space.

The incorporation of parking choice and PGI systems in general-purpose traffic models is considered by (Chatterjee and Hounsell 1999), with specific reference to the dynamic traffic model RGCONTRAM. The authors show how parking-related movements and the associated times and costs can be represented as special links in a traffic network model. They discuss the application of a travel choice simulator to investigate joint route and parking facility choice with and without PGI messages. They describe simulator experiments that varied parking prices, expected risk of waiting to park, waiting times and PGI messages, but do not present specific model specifications and estimation results. However, regardless of the particular form of a parking choice model, it is clear from the author's discussion how an information-based traffic model (i.e., one that allows en route path diversions based on messages) could represent and integrate parking information and choice as well.

Since parking search traffic is a poorly-understood but potentially significant component of city center traffic, it would seem that further research on driver choice of parking facility and driver response to PGI messages would be fully justified. Research results could be incorporated in information-based traffic models without requiring extensive modifications. With relatively little lead time after the research results became available, the resulting model systems could be applied to the practical analysis of parking search traffic and its impacts, and ultimately to the design of PGI systems.

3.2.3 SPECIFIC SYSTEMS AND EXAMPLES

This section discusses a few specific examples of driver response to information. The examples were chosen either because of their intrinsic interest, or because a considerable amount of information is available about them, thus allowing a more detailed discussion than was usually possible in the preceding section.

3.2.3.1 Variable message signs

Variable message signs (VMS) have been widely installed for freeway traffic management in most metropolitan areas. VMS are electronic message boards located in the close proximity to roadways. They represent a cost-effective mechanism to display short real-time messages to drivers approaching them. Of course, their effectiveness in real-time traffic operations is highly dependent on user response to the displayed information. A compounding factor is that, unlike an in-vehicle navigation system that can provide personalized routing information, VMS are constrained to display generic information to all nearby drivers. It follows that seemingly minor details of the displayed message may have a considerable impact on system performance.

This provides motivation to study the relationship between VMS messages and user response. A few studies have investigated this relationship. (Peeta, Ramos et al. 2000) examined the effect of different message contents on driver response under VMS. The issue was addressed through an on-site stated preference (SP) user survey. Logit models were developed for drivers' diversion decisions. The analysis suggested that content and level of detail of relevant information are factors that significantly affect drivers' willingness to divert. Other significant factors included socioeconomic characteristics, network spatial knowledge, and confidence in the displayed information. Results also indicated differences in the response attitudes of semi-trailer truck drivers compared to other travelers. These results provide substantive insights for the design and operation of VMS-based information systems.

A somewhat similar study was performed by (Wardman, Bonsall et al. 1997), also using a stated preference approach to undertake a detailed assessment of the effect on drivers' route choice of information provided by a variable message sign (VMS). Although drivers' response to VMS information will vary according to the availability of alternative routes and the extent to which the routes are close substitutes, the research findings showed that route choice can be strongly influenced by the provision of information about traffic conditions ahead. This has important implications for the use of VMS systems as part of comprehensive traffic management and control systems. The principal findings were that the impact of VMS information depends on: the content of the message, such as the cause of delay and its extent; local circumstances, such as relative journey times in normal conditions; and drivers' characteristics, such as their age, sex and previous network knowledge. The impact of qualitative indicators, visible queues and delays were examined. It was found that not only is delay time more highly valued than normal travel time (which is to be expected) but also that drivers become more sensitive to delay time as delay times increased across the range presented.

Most disaggregate-level studies of drivers’ response to VMS use stated preference (SP) survey data rather than actual traffic data. It is generally not possible to infer from traffic measurements the effects of a VMS on individual driver behavior, since the drivers’ intentions prior to receiving the messages are not usually known. One study encountered during the literature survey used aggregate traffic data. (Yim and Ygnace 1996) used loop detector data from the *Système d'Information Routière Intelligible aux Usagers* (SIRIUS) information network in Paris to investigate the effects of VMS on link flows. Time-series traffic data were analyzed to measure changes in mean flow rates at a selected link. It was found that variable message signs influence drivers to choose less congested routes when the drivers are provided with real-time traffic information, and that a driver's decision to divert is closely associated with the information pertaining to the level of congestion. In the Paris region, drivers received prevailing queue length information from the VMS. According to the data analysis, a reported queue length of 3 km seems to be a threshold at which a significant number of drivers choose to divert to an alternative route.

3.2.3.2 Compliance with prescriptive guidance

The question of user compliance with ATIS messages arises when those messages consist of prescriptive recommendations about pre-trip departure time or route choices, or about en route path switching decisions. It is of considerable interest to understand the factors that influence whether or not a driver will follow the recommendation, both as a means towards better design of prescriptive ATIS messages (to ensure higher compliance), as well as to model more accurately the effects of such messages at the individual or the system level.

Given basic traffic data on travel times or other measures of network conditions, either descriptive or prescriptive messages could equally well be generated from them. However, it does not follow that drivers' responses to these two different types of messages would be identical. The format and content of the two types of messages would necessarily be different, and could well elicit different reactions from drivers.

Descriptive guidance is in some sense more "neutral", in that it simply conveys information about network conditions, which drivers will interpret as they wish and are able. In contrast to this, prescriptive guidance is a specific recommendation to do a particular thing; drivers may question whether the recommendation is based on sufficiently reliable data, on decision-making criteria consistent with their own, or indeed on a knowledge of the network equal to their own.

On the other hand, prescriptive guidance may potentially provide a traffic control center with a more direct influence over drivers' tripmaking decisions and so on network-level traffic conditions. A considerable amount of underlying traffic data may be efficiently synthesized in the form of a simple recommendation to drivers. Particularly under incident situations, a center may feel it appropriate to intervene aggressively in drivers' choice processes in order to minimize avoidable traffic impacts and to restore normal conditions as rapidly as possible.

It should be mentioned that the distinction between descriptive and prescriptive guidance messages is not an absolute one. Indeed, as will be seen below, there is considerable evidence that the most effective ATIS messages combine descriptive and prescriptive aspects: information that describes a traffic situation together with recommendations that suggest an appropriate reaction. The information explains or justifies the recommendation in some sense, and drivers are more likely to comply.

The most common source of prescriptive guidance currently in operation is variable message signs. These may be used to suggest routes to drivers based on broad destination locations ("take route XYZ for points north"). The limited space available for message display is a major constraint, and the messages must be carefully designed to be clear and understandable. (Bonsall and Palmer 1999) discuss various aspects of VMS message design, and (Summala and Hietamaki 1984) present an earlier study of factors influencing traffic sign effectiveness.

In-vehicle units have the possibility of making much more detailed and personalized recommendations, but are not yet in common use. An early prototype system of this type was Siemens' Ali-Scout system, which was used in West Berlin's LISB deployment (Bonsall and Joint 1991b) and also in Michigan's FAST-TRAK program. It is based around an in-vehicle device that provides a simple keypad for entering data, and outputs both visual (simple text and direction arrows) and audible (synthesized voice) messages to the user. System beacons are installed at key locations on the network; these both transmit data to the in-vehicle units, and receive from vehicles information on their recent travel times. At the beginning of a trip, when a user inputs his or her intended destination, Ali-Scout first indicates the general direction to follow based simply on compass direction. However, when the equipped vehicle passes a beacon, it receives real-time travel time information from which it can determine a minimum time path. The in-vehicle unit then provides detailed driving directions (direction to take at each intersection) until the vehicle arrives in the vicinity of the destination. At that point, the in-vehicle unit reverts to a compass-direction mode, since the density of beacons is not high enough for the system to be able to provide detailed local area directions.

It is difficult to obtain information on prescriptive guidance compliance rates from aggregate traffic measurements such as link volume counts. Determining whether a driver complied or not with a recommendation requires knowing what the driver's original intention was, and also depends on knowing whether a particular message is relevant to the driver's situation. Such information is not generally available at the aggregate level, although license plate survey methods and driver questionnaires have occasionally been successfully used for this purpose (Dudek, Weaver et al. 1978), (Richards, Stockton et al. 1978).

For this reason, most research on driver compliance behavior has been based on experiments with individual drivers using travel choice simulators. Travel choice simulators place experimental subjects in a decision-making situation and record their response. Travel choice simulators focus on decision-making related to travel behavior such as route and departure time choice. They are less elaborate than the (much more expensive) vehicle simulators that attempt to faithfully replicate all aspects of the driving experience. Rather, they provide only the key elements of a choice situation under study, with enough detail to establish the context and to motivate users to respond in a realistic fashion. A travel choice simulation experiment can be viewed as a kind of stated preference survey in which the hypothetical choice scenarios are presented in a rather realistic manner.

For their research in to VMS compliance, Bonsall and co-workers developed first the IGOR travel choice simulator (Bonsall and Parry 1990; Bonsall and Parry 1991) and then the more sophisticated VLADIMIR travel choice simulator (Bonsall, Firmin et al. 1997). Both of these were PC-based programs that allowed subjects to "drive" through a network from a given origin to a given destination, following a route of their choosing. During the "trip", the program displays information on local traffic conditions and, at decision points, may also provide ATIS messages. The user chooses how to proceed at each such decision point, and the program records each such decision along with data about the conditioning factors such as traffic conditions, messages displayed, and others. The experimenter can vary these factors from one run to another in order to investigate their effects on drivers' decisions. In VLADIMIR the display took the form of actual photographs of locations along the routes being driven, along with a simple sketch map of the nearby network, text describing traffic conditions and any ATIS messages, and basic information regarding the progress of the simulation (elapsed time, etc.)

After careful comparisons of driver choices in the simulator with actual decisions by the same drivers in comparable situations on the network, (Bonsall, Firmin et al. 1997) concluded that the simulator was able to replicate driver behavior with a high degree of fidelity.

Similarly, Mahmassani and co-workers (Chen and Mahmassani 1993; Srinivasan and Mahmassani 2000b) used a travel choice simulator interfaced to the Dynasmart mesoscopic traffic model. The model represents 20 minutes of peak period traffic in a freeway corridor carrying roughly 11,000 simulated vehicles on three parallel facilities with several opportunities to switch from one to another. Experimental subjects (possibly several at a time) make departure time and route choice and switching decisions. These decisions are taken into account by the traffic model, which computes the traffic conditions that result from them (as well as those of the many simulated vehicles). The ATIS messages provided to the subjects are derived from the computed decisions, and so are consistent with those decisions (rather than being exogenously specified.) Strictly speaking, the messages are descriptive rather than prescriptive: they indicate the travel time to the (unique) destination on each of the three main alternative routes. However, in the simple context studied, the minimum time route is clearly the recommended one; the other considerations mentioned above that might affect compliance behavior do not come into play.

Based primarily on the results of travel choice simulator experiments, a number of general conclusions about driver compliance with prescriptive guidance have been obtained. Examples of such general conclusions include:

- drivers will reject prescriptive messages that they do not find credible. Factors affecting message credibility include the extent to which it is corroborated by local evidence, visible to the driver, about the alternatives and their conditions; and the quality of advice previously (and particularly very recently) received from the system;

- compliance is strongly affected by the driver's familiarity with the network. For a given prescriptive message, the compliance by drivers familiar with the network is generally about 10% less than that by unfamiliar drivers;

- compliance is highest for messages that combine information and recommendations; next highest for those that provide information only; and lowest for those that make recommendations only;

- one minute of delay mentioned in a VMS message has the same effect, in terms of affecting path choice decisions, as 1.75 minutes of actual delay in driving time;

- compliance is higher for recommendations about an immediate action than for vaguer advice about actions in the future. A recommendation that refers to a nearby problem location is more likely to be followed than one that does not;

- drivers have a certain reluctance to switch to a new route from one that they are already following. The reluctance is greatest if the recommended route seems to follow an alignment significantly different from that of the current route;

- socio-economic characteristics of the driver are also important influences on compliance. Among these characteristics are gender, age, level of driving experience, and (as already mentioned) degree of familiarity with the network.

A number of attempts have been made to model compliance behavior. (Srinivasan and Mahmassani 2000b) consider route choice behavior as influenced by both compliance and inertia mechanisms. The inertia mechanism reflects a driver's reluctance to modify a decision already made, while the compliance mechanism reflects a driver's tendency to follow (or to reject) routing advice. They specify and estimate multinomial probit route choice models that include these two mechanisms as latent variables, and conclude that the effects are significant. Simpler route switching models often include a dummy variable that penalizes routes if they are different from the one currently followed.

(Bonsall and Palmer 1999) discuss more particularly the modeling of driver choice of exit link at an intersection when guidance is provided. They estimate a number of simple multinomial logit models that incorporate variables such as travel time, message specific indicators (e.g., mention of accident or of road works), alignment of the exit link relative to current path, and so on. These models are intended for use in traffic simulation systems to predict the probability that individual drivers will proceed via the different possible exit links.

3.2.3.3 Shopping trips

Most studies on ATIS have focused primarily on commuting trips, which go to a fixed destination, tend to be repetitive in nature and involve tripmakers who are familiar with the transportation network. But it also of considerable interest to examine traveler response to ATIS during non-commute trips, where travelers have some flexibility in terms of destination choice and may not be as familiar with the transportation system.

(Mahmassani, Huynh et al. 2001) and (Kraan, Mahmassani et al. 2000) examined behavioral responses of non-commuters under real-time information during shopping trips. Utilizing results from an interactive stated-preference internet-based survey, the authors developed discrete choice models to investigate factors that influence en-route switching to alternate destinations and alternate routes during such trips. The fundamental difficulty in modeling this phenomenon derives from the manner in which information is provided to assist trip-making. The information provided and resulting user choices are interdependent. That is, the choice set presented to a tripmaker at a particular decision point is predicated on his/her previous decisions. Conversely, a tripmaker's decision in turn alters his/her subsequent information and choice sets.

The authors specified a model structure that overcomes this difficulty. It explicitly captures the conditional nature of the decision process. The model that they developed provides insight on en-route diversions during the shopping trip together with the factors affecting these decisions, especially with regard to the role of real-time information.

3.2.3.4 Transit information systems

Transit information systems provide transit users with static information on service such as routes, schedules, transfers and fares. They may also offer real-time information such as the anticipated arrival time of the next bus or train, and individualized information such as the route to follow or the expected travel time of a particular trip the user intends to make. Ultimately, transit information systems may offer their users a full range of trip planning, ticketing and real-time information services, integrated across the range of public transport modes; the Transport Direct system, currently under development in the U.K., ((Lyons, Harman et al. 2001)) is an ambitious step in this direction. (Casey, Labell et al. 2000 Section 3) provide a useful summary of the North American state of the art in transit information systems as of the year 2000. It is fair to say that currently deployed systems still have very rudimentary capabilities compared to their ultimate potential.

Real-time transit information can reduce the anxiety that users feel due to uncertainty regarding the duration of their wait. More generally, it may improve the quality of service perceived by transit users and likely increase transit's retention of its current patrons. Furthermore, providing information may possibly change non-users' attitudes toward public transit, and entice more travelers to use public transit.

A relatively limited number of studies have been undertaken to investigate the usefulness of these systems in attracting new transit passengers and improving the level of service of existing passengers. (Abdel-Aty 2001) used ordered probit models to study the effect of Advanced Traveler Information Systems (ATIS) on transit ridership. A computer-aided telephone interview was conducted in two metropolitan areas in northern California. The survey included an innovative stated preference design to collect data that address the potential of advanced transit information systems. The study's main objectives are to investigate whether advanced transit information would increase the acceptance of transit, and to determine the types and levels of information that are desired by commuters. The survey included a customized procedure that presents realistic choice sets, including the respondent's preferred information items and realistic travel times. The results indicated a promising potential of advanced transit information in increasing the acceptance of transit as a commute mode. It also showed that the frequency of service, number of transfers, seat availability, walking time to the transit stop and fare information are among the significant information types that commuters desire. Commute time by transit, income, education, and whether the commuter is currently carpooling, were factors that contributed to the likelihood of using transit following information provision.

Although such transit information systems are assumed to be of benefit, methods for evaluating these benefits under various conditions are limited. (Mishalani, McCord et al. 2000) developed a methodology that focuses on the potential benefits of bus arrival information systems to passengers waiting at bus stops under various supply and demand characteristics. Transit bus operations and passenger arrivals are modeled as a stochastic system where the operator uses real-time bus location data to provide to waiting passengers bus arrival time information that maximizes passengers' utilities. Simulation results reveal how the value of such information systems depends on the type of real-time data available to the operator, on bus operations characteristics, and on demand patterns. Results indicated that while the first two influence the value of information to passengers, demand patterns do not have a significant impact.

3.2.3.5 ATIS for maintenance and protection of traffic around construction zones

It is natural to think of applying ATIS to help manage traffic in and around construction zones. Such zones can create significant traffic disruptions. Because of their temporary and changing nature, most travelers will not be able to learn by experience what “typical” conditions are or how to avoid the most impacted areas. It is logical to suppose that providing real-time information to drivers in such circumstances would produce real benefits both to individual drivers and to network traffic conditions overall.

Surprisingly, there are very few examples of the use of real-time traffic information systems for construction zone traffic management in the U.S., and very little literature on the subject.

(Kratofil 2001) provides a brief but useful review of relevant literature. Based on his literature review, he then proposes a framework for quantifying the benefits of ATIS in construction zone traffic management, applying for this purpose a standard breakdown of ITS impacts into a number of such as mobility, safety, etc., and distinguishing between impacts to drivers, to the implementing agency, and to the community at large. He compares a “with” and “without” ATIS situation for a specific interstate highway reconstruction project using this framework. In most cases, quantification of the impacts of ATIS relies on values (for example, accident rate reduction impacts) that were derived for situations other than construction zone traffic management. He concludes his paper with a recommendation for collection of traffic data before and during the operation of the ATIS for construction zone traffic management, in order to begin accumulating quantitative results that could be useful for future design and evaluation efforts involving such systems. He also recommends the execution of surveys to better understand people’s usage and valuation of information from ATIS.

There is clearly considerable scope for ATIS MPT applications. Very little definite knowledge is available regarding either the design and operation of such systems, or traveler response to them.

3.3 What kinds of information do users want? How much will they pay for it?

In 1991, (Green, Sarafin et al. 1991) discussed the results of a study by a panel of experts of features that should be in driver information systems by the year 2000. To determine this, features were evaluated on the basis of three objectives that had been set by USDOT: (1) their effect on accidents; (2) their impact on traffic conditions; and (3) their fulfillment of driver needs. The analysis considered a very broad range of possible functions including communications, entertainment, office capabilities, way-finding, vehicle status monitoring, display of traffic signs and signals inside the vehicle, road hazard alerts, and traffic information. For each such function, the experts considered a variety of possible features that might implement the function. (In the entertainment function, for example, the possible features considered were cassette/CD player, radio and television.) Each feature was then ranked according to its contribution towards the stated objectives.

The five highest-ranked features were crash site hazard notification, in-car display of external traffic control signals, information on traffic congestion, indication of the presence of multiple compounding hazards in a driving situation, and information about road construction activities. All of these features are components of what we would now call an Advanced Traveler Information System, although some are still more advanced than is anything that has been prototyped to date. Features considered in the study which were given some of the lowest priorities, such as cellular telephone communications capabilities and radar detectors, are by now commonplace.

Considerable work since that time has attempted to identify user preferences for travel information system features. In this context, the term "features" refers to the different *kinds* and *qualities* of messages that might be provided by a traveler information system. By kinds of messages is meant the nature of the data provided in the messages – information on travel times or delays, location of incidents, specific route recommendations, etc. By quality of messages is meant their usefulness as it might be judged by a user – how up to date they are (their currency), their accuracy, precision, network coverage, the degree to which the message relates to the traveler's individual situation, and so on.

User preferences are obtained from various kinds of traveler surveys. In some cases, survey respondents are simply asked to express an opinion about various possible features: to state whether a feature is desirable or not, or to indicate the strength of their desire for the feature on an ordinal scale. Other survey methods involve placing the respondent in (hypothetical) situations where they must state their preference between alternative features, and so illuminate his or her tradeoffs between the features. Survey results may be analyzed by computing simple descriptive statistics or by estimating some form of econometric model. A number of these were discussed in the preceding sections.

(Llaneras and Lerner 2000), in a recent study of this type, compared user response to and preference for “basic” and “enhanced” ATIS services in the context of en route decision making; he used travel choice simulation experiments for this purpose. In these experiments, basic ATIS services consisted of descriptive information on incidents and congestion levels, and qualitative estimates of travel delays; enhanced services included all the basic services, but added information on alternative routes, various details about incidents, and a map display showing real-time traffic conditions. By analyzing the effectiveness with which users were able to translate the information received into travel improvements, the authors concluded that the following types of information were most valuable: data on incident location, type and delay; data on queue lengths; and recommendations about alternative routes, with directions to them. The real-time map display of traffic conditions was the information most frequently referred to by drivers in the experiments; however, human factors questions remain unsolved regarding the best way to present such information with minimal interference to driving.

When the survey choice situation involves both information features and money, it becomes possible to estimate an implicit willingness to pay for the feature, defined (in a utility based model) as the negative ratio of the marginal utilities of the feature and of money.[1] It must be emphasized that, to date, very few travelers have ever paid any money to receive travel information.[2] Answers about money in stated preference surveys are frequently biased because respondents know that they will not actually have to pay anything, regardless of what they say. Therefore, conclusions about willingness to pay for travel information are fraught with uncertainty, and the numbers obtained from such surveys should be interpreted in relative rather than absolute terms.

(Wolinetz, Khattak et al. 2001) list six broad factors that they hypothesize may affect travelers' willingness to pay for information:

- uncertainty: if there is little variability in traffic conditions from trip to trip, these is little need for real time traffic information. Non-recurrent congestion increases travel time uncertainty. Recurrent congestion, even through it is relatively more predictable, also adds uncertainty. Both of these effects may increase with trip length;

- information awareness: travelers who are aware of available ATIS services are more likely to express a willingness to pay for future services;

[1] When features are defined in discrete terms (e.g., information coverage of freeways only, of freeways and arterials, or of all roads), the "marginal" utility of a feature is the difference in systematic utility between two successive levels of that feature.

[2] This is beginning to change with the increasing number of new vehicles that offer in-vehicle navigation devices as a purchase option. However, the types of information currently provided by these devices is not yet as high quality as that assumed in most ATIS stated preference surveys.

- access to information: individuals who are willing and able to access real-time information through communication or computing devices may be more likely to pay for ATIS services;

- information use: individuals who already receive travel information via phone, radio or other conventional sources may be more willing to pay for ATIS services;

- situational and contextual factors: such as trip purpose, departure or arrival time flexibility, trip chaining requirements, and many others;

- socio-economic factors: background variables such as age, gender, income and education may be important influences on the willingness to pay for ATIS.

Research at the University of Michigan Transportation Research Institute (UMTRI) (Wallace and Streff 1993) studied the stated rankings of different types of travel information by drivers on different kinds of trips (commute trips, trips in a familiar area and trips in an unfamiliar area). This research compiled descriptive statistics on respondents' rankings of the relevance of different types of information on the route choice decision. The researchers were particularly interested in the influence of the different information types in the en route decision to switch from one route to a different one. For commute trips and those in familiar areas, information on travel delays and travel time reliability on the original and alternate routes were ranked the most highly. For trips in unfamiliar areas, the availability of travel directions for the alternate route was a highly ranked consideration.

(Mehndiratta, Kemp et al. 1999a) (see also (Kemp and Lappin 1999)) surveyed drivers who had had significant experience with prototype in-vehicle navigation devices in three recent field operational tests. Drivers' preferences with regard to information update frequencies, network coverage and information personalization were investigated in a series of attitudinal and tradeoff questions. The survey results were analyzed in a number of ways, including by estimating logit-form models of preference probabilities as a function of information quality and price. In general, the authors found that the most basic improvements in information quality over currently-available sources (general radio traffic reports, for example) were highly valued, but that further information quality improvements exhibited a pattern of decreasing incremental utility.

Geographic coverage and update frequency were both important attributes; logit model coefficients for the minimal level of provision of both of these had approximately similar coefficients. With respect to geographic coverage, door-to-door coverage was perceived as having little or no incremental benefit compared to coverage of freeways and arterials. Similarly, information updates several times an hour were clearly preferred to static information, but the added value of nearly continuous updates was small to negligible. Personalized information provision was not highly valued.

Few respondents were indifferent to the type of guidance – prescriptive or descriptive – provided by the system; they strongly preferred either one or the other. A majority of all respondents preferred to receive descriptive information (delays), although about 20% preferred prescriptive route guidance. Where sample sizes were large enough to allow such investigation of gender-related effects, it was found that women were more likely than men to prefer prescriptive guidance.

Most respondents indicated some willingness to pay for real-time traffic information; few indicated that they would not pay anything. The estimated willingness to pay ranged from $8-$10/month in Seattle, from $28-$36/month in Chicago and from $8-$20/month in Boston, depending somewhat on the particular information types and qualities considered. These values are higher than what is generally expected from other, perhaps more informal, analyses of user willingness to pay.

(Wolinetz, Khattak et al. 2001) is another recent investigation into user preferences and willingness to pay for different types of travel information. The survey covered both automobile and transit users in the San Francisco Bay Area, and asked respondents to rank possible information features of a hypothetical traveler information system; it also included pricing questions. Survey analysis was based on the computation of descriptive statistics. The most desirable information content options were constant updates, alternate route information, in-car computer information, expected delay data and route time comparisons. Many respondents indicated a willingness to pay at least some positive amount for high-quality real-time traffic information. The majority of these people prefer to pay on a per-request basis (as opposed to a flat monthly subscription fee.) Most expressed a willingness to pay up to $1 per request.

(Tsai 1991) reports on the results of focus groups held with commercial vehicle operators (truckers and bus drivers) regarding their preferences for information about the highway environment: traffic and weather. Desirable features included in the traffic data were: information on traffic congestion, accidents, lane closures, bridge closures, construction updates, alternate routes, low bridges, road weight restrictions and legal truck routes. Truckers identified specific areas (generally around the largest metropolitan areas) where such information would be particularly useful. Weather information needs included: notice of adverse or severe weather conditions, fog conditions, and identification of areas experiencing black ice. However, the expressed willingness to pay for such information was quite low.

(Ng and Barfield 1997) report on surveys of ATIS feature requirements of both private and commercial vehicle operators. Alternate route information was highly valued by all these users. Respondents indicated that the main reasons for choosing an alternate route were accidents, traffic volume levels and road construction activities. Around half the private and commercial drivers cited the gain in time by rerouting as the reason for switching routes. Accuracy and currency were found to be the most important attributes of the information provided by an ATIS or CVO application. Because drivers' observations of traffic conditions play an important role in motivating a route switch, the authors suggest providing live displays of real-time traffic conditions as a component of a traffic information system. They also suggest providing information (either en route or post trip) that confirms and validates the decisions actually made by a driver, in order to build confidence in the use of ATIS.

Survey and analysis issues that arise in investigations of user preferences for possible travel information system features are addressed directly or indirectly in a number of references in the literature. (Ng, Barfield et al. 1997) provide a high-level overview of survey design and analysis methods that might be applicable to such investigations, and furnishes extensive details about survey design issues and their resolution in several case studies. (Mehndiratta, Kemp et al. 1999a) discuss a number of stated preference survey design and analysis issues, including the possible presence of response bias (respondents give positive answers thinking it will please the surveyor) and non-commitment bias (respondents overstate their willingness to pay because no money is actually committed by answering). The authors also investigated the econometric problem of correlated error terms in the response by a single person to multiple related questions. They addressed the problem by specifying and estimating random parameter logit models, but found that this computationally-intensive technique did not result in estimates significantly different from those obtained using simple logit models.

3.3.1 ATIS MESSAGE RELIABILITY

Reliability is a feature of particular prominence in analyses of ATIS message attributes. Generation of high accuracy ATIS messages is a challenging technical task, for a number of reasons. Measurements of traffic conditions on a network will generally be made using a limited number of data collection devices (traffic detectors, probe vehicles, cameras, etc.) The measurements will inevitably be imperfect (imprecise and inaccurate) for a variety of technical reasons. Information of particular interest, such as assessments of the severity and clearance time of incidents, may not even be available until after special personnel (police, traffic crews) physically reach the incident site. Data communications and processing limitations mean that traffic measurements cannot be instantaneously converted into meaningful traffic messages. From imperfect measurements of a limited number of variables processed at time intervals that are large compared to characteristic times of traffic dynamics, it will be difficult, to say the least, to obtain and maintain a detailed and up-to-date picture of prevailing traffic conditions.

Furthermore, data on prevailing conditions may not be an accurate basis for determining the conditions that a vehicle will actually encounter on a path. (Ben-Akiva, de Palma et al. 1996) show analytically that use of prevailing conditions for ATIS messages can lead to a worsening rather than an improvement in traffic conditions, and explore the sensitivity of ATIS messages to inaccuracies and imperfections in traffic conditions. (Chen and Mahmassani 1991) investigated, using a mesoscopic traffic simulator, the reliability of route guidance recommendations based on prevailing times. They compared minimum paths and path times based on prevailing times with the actual minimum paths and path times using true (i.e. time-varying) link times and concluded that real-time ATIS messages based on "information on currently prevailing link trip times, with no attempt to predict future travel time or traffic conditions, may not be very reliable, especially at high levels of market penetration." However, guidance based on predicted traffic conditions requires forecasts and models, which may not be particularly accurate, and involves large amounts of computation, which will add to the time delays of the provided information. Thus, predictive guidance, even if it has the theoretical possibility of better matching a driver's actual travel experience, may be constrained in its accuracy by practical and computational factors.

A number of studies of user preferences for ATIS features have included consideration of message accuracy, as has been seen. (Madanat, Yang et al. 1995) included drivers' perceptions of information reliability as a latent (unobservable) variable in a route switch model and found it to have both direct and indirect effects on the probability of switching in response to information; the indirect effect came through its influence on drivers' general attitudes towards route diversion (another latent variable in the model). (Hato, Taniguchi et al. 1995) developed stated choice models of route switching behavior in which the accuracy of reported travel times was explicitly varied in different choice situations, and found that the information accuracy level was a significant variable in determining switching probability.

(Kantowitz, Hanowski et al. 1997a; Kantowitz, Hanowski et al. 1997b) explicitly consider the question of how much inaccuracy ATIS users will tolerate. They pose the issue in terms of the relative strengths of drivers' self-confidence in their knowledge of traffic conditions, and their trust in the ATIS messages. The authors conducted experiments using a travel choice simulator in which information on link conditions (light or heavy traffic) was intentionally degraded. They considered situations in which either 73% or 41% of the links had correct information. (These numbers come from prior work by the authors on reliability issues in human factors. Of course, in some cases, the incorrect information is harmful –when driver chooses a heavily congested link because it is reported to have light traffic, for example –while in others the error may be relatively benign.) They found that when 73% of the link reports were accurate, drivers still took account of the messages; while when only 41% were accurate, drivers ignored them. Drivers did not use accurate information as effectively in the familiar setting as in the unfamiliar setting. Also, inaccurate traffic information was more harmful in a familiar setting. Thus, it would appear that drivers are tolerant of a certain amount of error in ATIS messages, although drivers familiar with an area will expect a higher degree of accuracy from the information system.

3.4 User benefits from ATIS

The economic benefits that an ATIS user derives from ATIS services are very closely tied to the user's response to ATIS and to his or her willingness to pay for ATIS information: they are all aspects of the same internal evaluation and decision-making process. The discussions in Sections 3.2 and 3.3 have covered many aspects of ATIS user benefit evaluation.

It has been seen that the spectrum of possible user responses to ATIS information is vast, ranging from relatively simple behavioral responses like route switching to complex responses such as re-arranging ones schedule of daily activities. This range exceeds the gamut of responses conventionally considered in transportation benefit evaluation exercises, and indicates that considerable care must be taken in thinking about and quantifying their benefits.

In conventional evaluation approaches, user benefits are usually computed as a change in consumer's surplus, defined as the total difference between what each user is willing to pay (in money or in time) for something and the amount actually paid. Willingness to pay is deduced from the travel demand curve, expressing the amount of travel that would be made at different cost or time levels. The evaluation thus assumes that user benefits are tied to travel cost or time reductions.

This assumption is unlikely to lead to a complete and comprehensive approach to evaluating ATIS-produced user benefits. For example, peoples' re-arrangement of their daily activity schedules may lead to more rather than less time being spent in travel, as they are able to carry out more activities because of more precise planning. If one were to ask such people if they were better off because of ATIS, they would reply affirmatively, even though they spend more time traveling: the benefits they derive from the additional things they do more than offsets the opportunity cost and disutility of the time spent traveling. If this were not true, they would not have re-arranged their schedule.

Special cases of ATIS-produced benefits can be distinguished, and may lead to simplified evaluation procedures when it is known what are the preponderant impacts of ATIS in a particular situation. In general, of course, it will not be possible to know *a priori* what the main impacts of an ATIS on user behavior are likely to be.

For example, if the only effect of an ATIS is to cause someone to switch routes, it might be reasonable to evaluate the ATIS user benefits via the resulting savings in travel time or cost. (If the user has confidence in the ATIS, an additional benefit may derive from the reassurance of having made an *informed* route switch, as opposed to the stress that could accompany an uninformed decision.) However, as noted above, there are indications that ATIS-produced reductions in travel times are likely to be small, and that the most common effects of pre-trip ATIS will be in terms of departure time rather than path choice changes.

If the only effect of the ATIS is to provide more precise estimates of the travel time between activities at two locations, and so allow the user to spend more time at either trip endpoint, then it might be reasonable to evaluate the ATIS user benefits via the benefits of pursuing those endpoint activities. In this way, an estimate of the benefits of improved travel time reliability could be obtained.

(Small, Noland et al. 1999) carried out and analyzed stated preference surveys investigating the value of travel time savings in congested conditions, and the value of travel time reliability, to travelers and freight carriers. They found that travelers definitely impute a monetary value to travel time reliability; however, this value can be entirely explained in terms of the early or late schedule delay costs at the destination (i.e., the cost to a traveler of arriving earlier or later than her intended arrival time). After the schedule delay costs were accounted for, no residual valuation of travel time reliability could be detected from the survey results. Similar results were found for freight carriers, although the conclusions were less statistically robust: travel time reliability had a value to freight carriers, but this value was entirely attributable to the costs of late arrival compared to a scheduled time.

Brand (1998) has proposed a more general user benefit evaluation method that returns to the original economics approach based on willingness to pay. However, instead of attempting to estimate willingness to pay from a conventional time- or cost-based demand curve, he suggests estimating it directly, using stated preference surveys of current or potential users of ATIS services. Such surveys can pose questions in which respondents trade off service attributes against cost and, properly conducted and analyzed, can provide reliable information on users' willingness to pay for different service attributes or for entire systems. A number of willingness to pay results from stated preference surveys were discussed in the preceding section.

By obtaining willingness to pay in this direct fashion, many of the complications of a model-based approach are avoided. There is no need, for example, to estimate how ATIS users might re-arrange their daily activity schedules and tripmaking behavior, and then to evaluate the travel and non-travel benefits and costs of the re-arrangement: the effects of such possible changes are already incorporated in the users' responses to the stated preference surveys. This user benefits estimation method has the potential of being both simpler and more accurate than adaptations of conventional transportation evaluation methods to the very different properties of ATIS, although the usual caveats regarding stated preference surveys continue to apply.

3.5 Day-to-day effects and learning

ATIS is a new and evolving set of technologies, and new ATIS users will need to learn about its features, capabilities and performance. While learning about and using ATIS, individuals will inevitably have a variety of experiences with it, both positive and negative. Over time, these experiences will in some way shape peoples' attitudes towards and use of ATIS. At a larger scale, the mechanisms by which people learn about ATIS and filter their experiences with it will strongly affect the public's overall acceptance or rejection of ATIS technologies.

Perhaps for these reasons, a number of researchers have investigated the day-to-day learning processes associated with ATIS. Indeed, it appears that this subject has already been more intensively investigated than learning processes associated with conventional traffic equilibrium.

(Iida, Akiyama et al. 1992) provide an example of a study of learning processes in a conventional equilibrium context. They analyzed the dynamics of route choice behavior in simulator-based experiments that asked the participants to respond to repeated hypothetical route choices. In the analysis, travelers depart from a single origin to a single destination connected by two parallel alternative routes. Day-to-day variations in traffic conditions are represented by route travel time changes. Travel time prediction errors (the difference between predicted and actual travel time) as well as actual travel times are treated as "experiences" accumulating through the experiments. It was found that assumptions about learning behavior strongly affected the day-to-day variability of traffic flow; however, none of the assumptions considered led to flow equilibrium. The authors conclude from this that existing traffic assignment models may not be adequate representations of actual traffic phenomena.

In the context of ATIS, (Iida, Uno et al. 1999) performed a study to identify changes in drivers' route choice mechanisms following the introduction of ATIS. They also investigated the influence of the accuracy guidance information on the route choice mechanism. The study used a travel choice simulator with which subjects repeatedly traveled between the same origin and the destination in the morning. During the experiment, the subjects learned about, and accumulated knowledge of, the network and information system. It was found that introduction of ATIS did change the decision mechanism that drivers applied, and that the quality of the provided information affected the nature and permanence of the change.

A similar study performed by (Vaughn, Abdel-Aty et al. 1993b) analyzed the accuracy of information provided in modeling drivers' sequential route choices. This study also used discrete choice modeling framework to model sequential route choices. Experimental sequential route choice data under the influence of ATIS was collected using a PC-based travel choice simulator. The experiment collected information on drivers' pre-trip route choice behavior at three levels of information accuracy: 60 percent, 75 percent and 90 percent. An analysis of variance was performed on the data to investigate the interrelationships among the different variables in an attempt to identify factors that significantly influence route choice behavior and learning. An attempt was made to model sequential route choice behavior using a binary logit model formulation; the results were mixed. It was assumed that drivers update their knowledge of the system based on their previous experiences; therefore an information updating function was specified and incorporated into the model. The results indicate that drivers can rapidly identify the accuracy level of information being provided and that they adjust their behavior accordingly. There is also evidence that indicates that an accuracy threshold level exists, below which drivers will not follow advice and above which drivers readily follow advice.

(van Berkum and van der Mede 1999) proposed a very general dynamic model of ATIS-guided route choice that includes behaviors based on perceived utility maximization, habitual choice and compliance with prescriptive guidance. Irrespective of the choice rule operating, individuals learn from their experiences. After each trip, the experienced travel time is used to update the mean expected travel time and the travel time variance for the chosen route. Descriptive and prescriptive guidance information influence route choices in different ways. Descriptive information may be incorporated into the perceived utility of alternatives for the subsequent choice. Prescriptive guidance can overrule the perceived utility maximization and habitual choice behaviors. The degree to which guidance affects the decision depends on the credibility of the information, and the credibility is influenced in turn by previous experiences with the information system.

In modeling dynamics, it is necessary to observe the behavior of a decision-maker over time. Investigating route switching in a dynamic context enables the calibration and testing of richer model specifications by incorporating repeated measurements, heterogeneity, within-day and day-to-day influences of variables, and state dependence effects. The multinomial probit framework (MNP), though well suited to tackle these challenges in dynamic models with a few periods, is prohibitively expensive for panels of longer duration.

To address the needs of modeling dynamic route switching over a large number of decision periods, (Srinivasan and Mahmassani 2000a) proposed a dynamic kernel logit model that retains the flexibility of multinomial probit while exploiting to some extent the computational tractability of the logit model. They applied the model to analyze the influence of systematic effects on route-switching behavior under ATIS. The effect of trip maker characteristics, trip characteristics and traffic conditions, experiences in traffic, and attributes of ATIS information are examined in this context. They also investigated heterogeneity effects in route switching behavior. Finally, time-dependent effects in route switching behavior are examined in two ways. First, at the systematic level, the influence of past experiences on current behavior is assessed. Second, dynamic effects were investigated via the structure of the utility disturbance terms. At the unobserved level, time dependence effects are examined by specifying suitable variance components. The variance-covariance structures are tested for the presence of temporal correlation (both within day and day-to-day), in addition to serial correlation (due to repeated measurements).

Many analyses of driver-network transportation systems assume that the systems are in equilibrium. Equilibrium analyses presuppose that the driver is rational and homogeneous, and has perfect information. (Nakayama, Kitamura et al. 2001) suppose, on the contrary, that people have cognitive limitations. A driver is assumed in this study to adopt simple rules when choosing a route. The authors develop a simulation system in which drivers' learning is simulated through a genetic algorithm that, over time, that generates and modifies a set of route choice decision rules. The results of simulation analyses can be summarized as follows: Drivers do not become homogeneous and rational as equilibrium analyses presuppose; rather, there are less rational drivers even after a long process of learning, and heterogeneous drivers make up the system. Drivers' attitude toward and perceptions of each route do not become homogeneous either, but become bipolar. The results point to the need for a critical appraisal of the foundation of the equilibrium analysis of network flow.

(Ozbay, Datta et al. 2001) proposed the use stochastic learning automata (SLA) to analyze drivers' day-to-day route choice behavior. This model addresses the learning behavior of travelers based on experienced travel time and day-to-day learning. In order to calibrate the penalties of the model, an Internet based Route Choice Simulator (IRCS) was developed. The IRCS is a traffic simulation model that represents within day and day-to-day fluctuations in traffic and was developed using Java programming. The calibrated SLA model was then applied to a simple transportation network to test if global user equilibrium, instantaneous equilibrium, and driver learning have occurred over a period of time. It was observed that the developed stochastic learning model accurately depicts the day-to-day learning behavior of travelers. Finally, it is shown that the sample network converges to equilibrium, both in terms of global user and instantaneous equilibrium.

While many travel behavior studies that deal with day-to-day learning have focused on modeling route choice behavior under information, fewer have examined day-to-day processes in departure time choice behavior with ATIS. The motivation in modeling departure time choice dynamics stems from the following considerations. The departure time decisions of commuters on a given day significantly influence the within-day distribution of traffic, congestion and queuing patterns on the network in the peak period. Accurate models of departure time adjustments can translate into a robust time-dependent OD prediction capability that is an essential component for dynamic traffic modeling and assignment techniques. In addition, since departure time variations influence the network flow evolution from day-to-day, models of departure time choice dynamics are important for characterizing and analyzing dynamic network states and the associated costs. Dynamic models of departure time choice play an important role in demand forecasting, as an integral component of activity-based demand modeling framework.

(Mahmassani and Chang 1986) performed an exploratory analysis that included 1) the explicit treatment of the day-to-day dynamics of departure time decisions, 2) the specifications of mechanisms by which individual users adjust their decisions on a daily basis, given prior experience, 3) the boundedly-rational heuristics that are assumed to govern individual tripmakers' behavior, and their use in a modeling framework that recognizes the interaction between user behavior and system performance, and 4) the use of a special-purpose traffic simulation model to study the dynamics of user behavior. An extension of this work was conducted by (Mahmassani and Stephan 1988) in two directions: 1) the inclusion of the route choice dimension in addition to that of departure time and 2) the consideration of two user groups with different information availability levels interacting in the same simulated commuting system. The effect of information availability on the behavior and performance of given user group was of particular interest. In this regard, the results of this experiment are broadly consistent with a priori expectations; that is, users with more information clearly outperform those with limited information when both are competing in the same system. The interdependence between route choice and departure time decisions is another important aspect of user behavior addressed in this paper. The exploratory aggregate analysis considered here points to the precedence of departure time shifts over route shifting in dealing with experienced unpredicted congestion in the system.

The above mentioned works in day-to-day departure time choice modeling do not propose specific models of the departure time adjustment process. (Srinivasan and Mahmassani 2001) addressed this by investigating alternative mechanisms commuters' day-to-day departure time adjustment behavior. The mechanisms they considered include: utility maximization from unordered alternatives; ordinal response mechanism (where thresholds are corresponding to choice alternatives are ordered); sequential greedy search process; and a two-stage nested adjustment process. Econometric models are proposed corresponding to these mechanisms and implemented using departure time adjustment data obtained from interactive simulator-based experiments. The results indicate that the observed departure time choice dynamics is consistent with a sequential greedy search process. Under this mechanism, users continue to search for acceptable adjustment alternatives in a sequential and ordered fashion, until a satisfactory departure time choice is obtained. The results also indicate that network conditions, users' past experiences in the short and longer-term, and the nature and type of real-time information supplied by ATIS significantly influence the adjustment behavior of commuters. The models and results have significant applications in demand forecasting, network state prediction, and the evaluation of transportation control measures.

All of the above studies considered single-purpose trips from origin to destination. In fact, many trips involve multiple purposes and intermediate stops; this is called trip chaining. Trip chaining can significantly impact travelers’ route and departure time switching behavior. Trips with intermediate stops are more likely to involve switching than trips without stops. (Mahmassani, Hatcher et al. 1991) addressed the daily variation of trip-chaining behavior of commuters, and related it to various attributes of the commuter, the workplace, and the commute. The paper addresses the day-to-day variation of three key aspects of the home-to-work commute: 1) the time of departure from home; 2) the frequency, purpose, and duration of intervening stops between home and work; and 3) the path actually followed through the network. It is based on two-week detailed diaries of actual commuting trips completed by a sample of auto commuters in Austin, Texas. About 25 percent of all reported commutes contained at least one non-work intermediate stop, underscoring the importance of trip chaining in commuting behavior. These multipurpose trips are shown to influence significantly the departure time and route-switching behavior of commuters.

Although considerable attention has been given to incorporate day-to-day learning in route and departure time choice modeling, the same cannot be said about modeling mode choice. The only work that was found during the course of the literature review is by (Aarts, Verplanken et al. 1997). This study focuses on travel mode choice behavior in order to test theoretical propositions as to habitual decision making. It investigates the effects of habit on information processing during judgments of travel mode use. The study used multiple regression analysis to test the hypothesis that habit is negatively related to the elaborateness of information processing preceding judgments of travel mode use. The study focused on the judgment of bicycle use for short distance trips. It is expected that individuals who have developed a strong bicycle choice habit apply less elaborate information processing strategies compared to those who have not developed such a habit.

3.6 Human factors issues

A driver's ability to navigate through a complex environment is largely dependent on the type and extent of cognitive structures representing that environment, the goals of the driver, and the ability of the driver to stay oriented. These three areas, founded in psychology and environmental cognition, are functionally related. First, a destination and travel plan must be formed. Second, knowledge of the local or global network must be known or acquired. Finally, a reference system must exist to relate the driver to the environment. The cognitive map has been hypothesized as the basis for mentally storing or representing information about the physical world. The internal format of remembering this information could have profound effects on the ease with which one can assimilate information presented by an Advanced Traveler Information System (ATIS). If the information is mentally stored in a prepositional format, then specific verbal directions may be desirable. However, if the information if the information is in a format analogous to the real world, a different representation, the map for example, may be desired. In addition, the spatial and verbal skills of drivers may vary significantly among individuals; thereby influencing their ability to use different navigational display formats. Human factor issues of concern include the format and coding of navigation system information, the attentional demand and safety issues of displays and controls, and agreement on general guidelines for the development and manufacture of ATIS.

A number of research studies have been reviewed that deal with human factors involvement in the design and use of ATIS. Some of them deal with the application of human factors guidelines and design decision aids for ATIS and ATIS displays. The questions that the designers must answer when developing displays for ATIS, which will affect or have an impact on both the safety and usability of the system are: (i) What information should be included in the ATIS that is being developed? (ii) What functions of the ATIS should the driver be allowed to use? (iii) To which sensory modality (e.g., auditory, visual, tactile) should information items be allocated? (iv) What format (e.g., text, map, tone, voice) should be used to present the information?

(Mollenhauer, Hulse et al. 1997) explored the decisions that designers must make when developing ATIS displays. They described a design support process that has been developed to help formulate answers that reflect current human factors research and accepted design principles. Examples of decision tools that make up this process are provided along with a description of how these tools can be used together to aid in the design process. To analyze the information format options, "trade study" analysis is used to aid in design decisions. These analyses serve as systematic aids for complex decision making. In addition, specific results are also presented and discussed.

(Landau, Hanley et al. 1997) reviewed the following topics for guideline availability and applicability to an ATIS:

- input methodology: The design of the input mechanisms for an in-vehicle system must consider the accuracy and speed required for transactions;

- display and information characteristics: The research covers guidelines related to both legibility and readability of a display;

- auditory display characteristics: Auditory displays include both nonverbal and verbal aural displays. Nonverbal displays use auditory alerting signals to signify events. Verbal displays use voice signals or messages to signify events and to provide more complex information. Auditory displays can supplement visual systems;

- human-computer interaction: The interaction between a driver and an ATIS system will be modeled to a great degree on human-computer systems because the nature and complexity of the transactions are so similar to current computer interfaces. Therefore, the applicability of human-computer interface guidelines is reviewed; and

- navigation information format: Navigation information is typically portrayed by maps that provide direction and distance relationships in a plan view presentation. Another type of navigational format is turn-by-turn sequential list.

The successful implementation of ATIS depends on user acceptance of its products and services. Information on user acceptance could be applied to the design of ITS products and services, as well as to the development of ATIS implementation strategy. User acceptance is particularly important to the successful implementation of ATIS because the accuracy of traffic information it conveys is dependent on the number of ATIS equipped vehicles. Receiving inaccurate information from an ATIS device may break the trust the driver has in the system and lead to user rejection. Consumer rejection of ATIS, in turn, may lead to decreased system reliability and accuracy. ATIS is unique in that the degree of consumer use affects system effectiveness. Thus, to optimize ATIS accuracy, initial acceptance of ATIS should be maximized.

The results of the study done by (Wochinger and Boehm-Davis 1997) indicate that the drivers showed strong differences in their initial preferences for maps and text directions. However, most of the participants rated ATIS higher than the other aids after a “hands on” experience with it. Older drivers in particular may be unlikely to embrace a technologically innovative system. An ATIS implementation strategy can facilitate user acceptance by presenting information to positively influence customer reaction to ATIS.

Giving drivers advance warning of an event can affect route choice and safety related factors such as driving speed. However, the success of such systems depends largely on the ability of drivers to assimilate, retain and act on the information received. These processes rely on the application of ergonomics to the design of the system's man-machine interface (MMI). So, it is very important to know how drivers assimilate information and retain it over time. (Graham and Mitchell 1997) carried out a road based experiment to examine both the assimilation process and the retention of information over time. Measures of recall performance and eye glance behavior were used to assess three factors associated with the design of driver information systems: the length of messages, the timing of messages, and driver age. The study compared the performance of two age groups of drivers using the system. Recommendations were made concerning the amount of information that should be displayed on the screen, the timing of messages in relation to events, and the presentation of message screens. (Akamatsu, Yoshioka et al. 1997) conducted field experiments to explore driver behavior and the processing of information when navigation systems are used in real urban areas. Driver behavior while using a navigation system in the central area of Tokyo was recorded by means of small video cameras, and the landmark information used by drivers was analyzed using the "thinking aloud" method. In the analysis, verbalized words were categorized into several types of landmark information.

Another important consideration is the amount of driver workload that is involved in using in-vehicle navigation or route guidance system. It is very important to know how the characteristics of route guidance systems affect the attentional demand and efficiency of the driving task and to understand how drivers react to complex route guidance systems under varying task demands resulting from driving in different types of roads. (Srinivasan and Jovanis 1997) used a high fidelity driving simulator to collect detailed driving performance data in an investigation of the following questions:

- do electronic route guidance devices lead to better driving performance compared to paper maps?
- do audio route guidance systems lead to better driving performance and lower workload compared to their visual counterparts and paper maps?
- does a head-up turn-by-turn display in combination with a head-down electronic route map lead to better driving performance and lower workload compared to a head-down electronic route map?

Its also important to know about drivers' route choice behavior in the presence of ATIS from human factors perspective. It might involve knowing drivers' behavior in the presence of different forms of information. (Katsikopoulos, Duse-Anthony et al. 2000) studied drivers' route choice behavior when travel time information is provided under varying degrees of cognitive load. In this study, travel time variability is presented by giving drivers a range of possible travel times for routes with an uncertain travel time. A route (main) with a certain travel time and a route (alternate) that could take a range of travel times are described. This study investigates the effects of average travel time and travel time variability. Scenarios were considered in which the average travel time of the alternative route was smaller than, equal to, and greater than the certain travel time of the main route. Attempts were made to determine whether the effect of range is a function not only of framing but also of the cost of being late. This research also tests whether participants make the same choices while driving as they do when sitting still.

4 NETWORK IMPACTS OF ATIS

4.1 From individual- to network-level impacts

Boyce (1988) speculated over a decade ago that as tripmakers begin to experience the benefits of better travel information and decisions from ATIS, they would come to re-consider and adjust many of their significant life decisions, including where they live and work, and how they arrange their daily activity schedules. Large-scale changes in residential and employment locations would inevitably lead to major shifts in urban, suburban and exurban land use and structure, affecting in turn the spatial pattern of transportation demand. Rearrangements in daily activity patterns – consolidating trips in to chains or splitting chains into individual trips, making trips at different times – would affect the temporal pattern of transportation demand. ATIS and related ITS technologies, he argued, were not just traffic information and management tools, but had the potential to affect travel demand at a fundamental level.

Clearly, significant rearrangements of the basic organization of peoples' activities, and the resulting changes in travel patterns, would have repercussions well beyond the transport sector: environmental and energy consumption changes, either positive or negative, would also follow, to cite only the most obvious.

Few analyses conducted to date have taken such a broad view of network-level ATIS impacts. Part of the reason for this, no doubt, is that there is very little empirical basis for quantifying the nature and magnitude of some of the effects identified by Boyce, Brand and others. In any case, most analyses are conducted within a short-term analysis time frame, and consider primarily the network effects resulting from route choice (occasionally route and departure time choice) adjustments by tripmakers in response to ATIS.

Clearly, the magnitude of network-level ATIS impacts depends on the number of drivers receiving information. If only a few drivers obtain guidance, they may benefit from improved decision-making, as discussed in Section 3.4, but any tripmaking changes they may make will have negligible impact on network conditions.

As more drivers receive ATIS messages, the aggregate effect of their reactions to it becomes important; indeed, because of the congestion externality, the aggregate effect may be out of proportion to the magnitudes of the individual reactions. This aggregate effect depends both on what particular guidance messages are disseminated (including where, when and to whom) as well as on how drivers react to the messages.

(Ben-Akiva, de Palma et al. 1991) identified some of the possible adverse network-level effects that can result from guidance dissemination. These include:[3]

- overreaction, which occurs when a significant number of drivers receive identical messages and react in roughly the same ways. This could cause congestion to transfer from one route to another or even produce oscillations in path flows; and

- concentration, which occurs when driver information reduces the natural variability of individual drivers' decisions and leads them to act similarly, possibly leading to congestion increases.

In either case, the distinct possibility exists that providing guidance messages could worsen rather than improve traffic conditions.

This means that as the number of drivers receiving guidance (the *market penetration rate*) increases, it becomes important when generating guidance to take account of the effects of the guidance itself on drivers and traffic conditions. This is necessary not only to avoid particular effects such as overreaction and concentration, but more generally to ensure that the guidance messages that are disseminated based on traffic conditions remain consistent with those conditions after drivers receive the messages and react to them.

Evaluations of network-level ATIS impacts can be categorized as theoretical (model-based) or empirical. Many, but not all, of these have focused on travel time and its variability, the measures most immediately impacted by ATIS. Examples of more general approaches to theoretical or empirical ATIS impact analysis follow.

(Thill and Rogova 2001) describe a sketch-planning model tool to screen proposed infrastructure-based ATIS projects (such as VMS) based on their improvements in travel delays, traffic safety and environmental quality. For VMS evaluation, for example, it assumes a basic corridor topology consisting of a main route and a diversion alternative. In case of an incident on the main route, it computes total time as the sum of time traversing the main route, time spent

---

[3] They also identified the individual-level problem of oversaturation, which occurs when the amount of information a person receives is too great to be effectively processed into a rational decision in the time available to make the decision. The general problem of information oversaturation is exacerbated in a driving context because of the accident-producing potential of driver distraction or confusion. In practice, the need to avoid oversaturation limits the amount and complexity of information that can be conveyed in messages to drivers.

queuing there, time spent traversing the diversion route, and time lost when the diverted traffic merges with mainline traffic. Delay reductions are calculated by comparing total time in baseline and deployed situations. The basic traffic characteristics determined during the travel time calculations are used as inputs to the safety and environmental benefits estimates. Safety benefits (reductions in primary and secondary accidents, distinguished in terms of fatalities, injuries and property damage only) depend on the amount of congestion on the main and alternative routes. Environmental benefits (reductions in VOC, CO, NOx and fuel consumption) depend on calculated VMT and speeds. The evaluation tool utilizes default values for key parameters such as diversion rates, accident rates, and emissions factors, rather than computing them endogenously.

Brand (1998) (see also Brand 1995) argued that evaluation of ATIS economic impacts at the disaggregate level needs to take account of the many adjustments in the individual, household and business activities that would come about from improved travel information. These adjustments could generate considerable utility or economic benefits even if more trips were made and more time spent traveling. The nature of these benefits could best be determined by individual-level investigations (stated preference surveys in which ATIS users traded off alternative possible system features against possible costs), and the most accurate determination of system-level ATIS impacts would come from aggregating such individual level results. Indeed, he argues that considering only the total travel time or VMT impacts of ATIS might lead one to seriously erroneous conclusions regarding its benefits.

(Arnott, de Palma et al. 1991) also caution against simplistic measures of ATIS benefits. They argue that, since congestion is an un-internalized externality, drivers' reactions to information about congestion may not be efficient (in the economic sense). The reactions may increase rather than decrease congestion. The authors provide a simple model, involving route and departure time choice in a two-route corridor with stochastic capacities. They show that average travel costs are reduced when drivers receive perfect information about the route capacities; however, with imperfect information an un-internalized congestion, drivers may change their departure times in a way that worsens congestion.

Related points are made by (Emmerink, Nijkamp et al. 1994). ATIS may enable drivers to avoid excess travel from uninformed path choice decisions or parking search; on the other hand, because of the improved efficiency of travel, and the activity adjustments discussed in Brand (1998), ATIS may induce more demand for travel. They note that an ATIS may involve both positive and negative externalities: an additional ATIS-equipped driver will generally increase the travel times experienced by other equipped drivers, but will decrease the times experienced by unequipped drivers. The existence of these externalities may lead to market failure if not corrected. The authors suggest a combined system involving both ATIS and road pricing both to internalize the ATIS externalities and more generally to combat congestion.

The following section describes a number of selected model-based analyses of network-level ATIS impacts; as mentioned, most of them use travel time or related quantities as the principal measure of performance.

4.2 Conclusions from computational and analytical models

Model-based analysis of ATIS impacts may involve analytical (purely mathematical) or computational methods. In principle, analytical methods can provide exact solutions, but sometimes a problem has to be simplified to enable a solution to be determined. Computational methods allow greater latitude in representing problem features and assumptions, but it is sometimes difficult to derive general conclusions from the solutions obtained to particular problems or, sometimes, to know if a correct solution to a problem has been obtained at all.

(Al-Deek, Martello et al. 1989) used standard traffic simulation packages to determine travel times in recurrent and non-recurrent (incident) congestion conditions in a portion of the SMART corridor in Los Angeles, California. They also surveyed commuters in the corridor to determine their usual and diversion routes. The authors then compared path times on the usual and "optimal" (minimum travel time) paths for a number of OD pairs. The reasoning was that, under perfect information from an ATIS, drivers would pick the minimum travel time path. The analysis thus indicates the magnitude of travel time benefits that ATIS might produce in this corridor.

The results indicated that under the recurring congestion scenario, the travel time savings from utilizing the shortest path were generally negligible (less than 3 minutes for a 20-25 minute trip) compared to the travel time on other paths (usually the freeway-biased path). Under the incident congestion scenario, travel time savings from choosing the minimum time path were found to be significant (greater than 3 minutes) during certain times in the analysis time frame. The greatest time savings accrued during the time slices immediately following the incident occurrence, with a maximum savings of 10 minutes for a 30 minute trip.

Note that the analysis does not take account of the effects of the guidance itself on drivers' decisions and the resulting traffic conditions. If a significant number of drivers switched paths in response to traffic information, travel times would no longer be the same as those used in the path time calculations and the "minimum" path found on the basis of the original times might no longer be so. This suggests that the estimated savings are upper bounds.

(Koutsopoulos and Lotan 1989) and (Hamerslag and van Berkum 1991) carried out studies of ATIS travel time impacts using a static stochastic user equilibrium assignment model. Informed and uninformed drivers were distinguished in terms of their travel time perception errors, modeled via the standard distribution of the error term in the probit-based path choice model. Informed drivers had small (possibly zero) error term standard deviations, signifying that their perceptions of travel times were close to reality; uninformed drivers were the opposite, and so, based on their inaccurate perceptions, might make path choices that were significantly sub-optimal.

In (Koutsopoulos and Lotan 1989), path choice was the only travel decision impacted by ATIS. They applied the model to a small urban area. They found that the difference in average travel times between informed and uninformed drivers (measuring the value of the information to the informed drivers) decreased as the network congestion level increased; however, informed users always had lower average travel times than uninformed users. With increasing percentages of informed users, the average travel time of both informed and uninformed users increased somewhat. Overall, however, the weighted average travel time decreased monotonically (but not always linearly) with increasing percentages of informed users.

(Hamerslag and van Berkum 1991) generalized the approach somewhat to allow trip distribution to depend on travel time perceptions. The authors used a combined static distribution-assignment model to predict the trip distributions and network traffic conditions that would result from different levels of information accuracy. The authors also considered a variety of networks. This is one of the few published quantitative studies of the possible impacts of ATIS on overall trip patterns (as opposed to path and departure time choice).

It was found that in all cases the total amount of travel (vehicle-kilometers of travel or VKT) decreased with decreases in the level of travel time uncertainty, as the spatial distribution of trips adjusted to the improved perception accuracy. The authors concluded that an ATIS might reduce VKT in urban networks by 15—20 percent and in regional networks by 5—10 percent.

The authors note that static traffic models that represent ATIS indirectly via its reduction in perception errors are not able to (i) analyze traffic dynamics at short or medium time scales or (ii) analyze specific ATIS characteristics or features.

(Al-Deek and Kanafani 1993) present an analytical queuing model of an idealized corridor with two parallel routes. An incident occurs on the main route. An ATIS diverts equipped vehicles to the alternate route in a way that maintains user-optimal travel times, based on moving and queuing, on the two.

The study results show that, following an incident, guided traffic is better off than unguided traffic during the diversion period that precedes the establishment of a travel time equilibrium between the main and diversion routes. However, this advantage is substantially reduced when a queue forms on the alternate route. The benefits to guided traffic are insensitive to the fraction of vehicles equipped with ATIS as long as this fraction is below the critical value that causes a queue to form on the alternate route.

When the alternate route is congested, the benefits to guided traffic become sensitive to the fraction of vehicles equipped with ATIS. The benefits to guided traffic decrease while the benefits to unguided traffic increase with this fraction. Thus, as the proportion of guided traffic increases, the difference in benefits between guided and unguided traffic narrows. System benefits increase proportionally with the market penetration rate as long as it is below the critical fraction, but increase less than proportionally when a queue forms on the alternate route.

(Emmerink, Axhausen et al. 1995) carried out studies of the travel time impacts of ATIS in a small network subject to random incidents, using a stochastic discrete-event simulator. Drivers were assumed to be boundedly rational, meaning that they only revise their current path if information that they receive about expected path times indicates the opportunity for a significant travel time gain. The authors investigated the travel time impacts of a number of ATIS parameters, including the market penetration rate and the information update frequency. The latter parameter determined how often updated estimates of remaining travel time to the destination (based on continuously changing travel conditions) were disseminated to drivers: periods of 1, 5 and 10 minutes were considered. Drivers receiving such information were assumed to combine it with their own prior experience to form their individual estimate, which was then the basis of a boundedly rational path switch decision.

The authors found that network-wide travel time decreased with increases in the market penetration rate. It was also found that the additional benefit to equipped drivers decreases quickly as the level of market penetration increases. Non-equipped drivers are also affected by the presence of equipped drivers, and their travel time benefits depend upon the level of market penetration as well. A decrease in the updating frequency has an adverse effect on network-wide performance. The size of this negative effect depends on the market penetration rate. However, the network-wide situation at full market penetration is still considerably better than without information.

(Mahmassani and Jayakrishnan 1991; Mahmassani and Peeta 1993) describe the Dynasmart simulation-assignment model developed at the University of Texas at Austin. Dynasmart is a mesoscopic traffic simulator, meaning that it simulates the movement of individual vehicles moving through a network in accordance with macroscopic flow rules (e.g., speed-density relationships.) It simulates several different route choice rules including dynamic system optimality, dynamic user optimality, and a bounded rationality rule in which drivers receiving en route information about path conditions will only switch paths if the expected improvement exceeds a threshold amount. Dynasmart has been widely used for investigations of route guidance. DynaMIT (Bottom, Ben-Akiva et al. 1999) is another mesoscopic traffic simulation model that is explicitly designed for route guidance applications.

(Hall 1996) reviews a number of simulation studies of ATIS total network-level travel time reduction benefits as a function of the ATIS market penetration rate. A number of these studies have found an “inverse U” shaped relationship, with maximum total benefits typically occurring at market penetration rates of 20—30 percent. Some have found negative benefits (i.e., increases in average or total network times) at high market penetration rates.

Using a simple analytical queuing model somewhat similar to the one applied by (Al-Deek and Kanafani 1993), he shows that for some network structures increasing the market penetration of accurate (i.e., experienced travel time) information cannot result in an increase in total network travel time; however, increasing the provision of instantaneous time estimates might in fact result in such an increase. He speculates that some of the results reported in earlier simulation and analytical studies may be due to use of instantaneous rather than experienced travel times in the models applied by their authors; this would lead travelers towards dis-equilibrium behaviors and produce disbenefits.[4] He argues that, in any case, the determination of the optimal market penetration rate is an irrelevant issue, since the rate should be determined through market forces and not enforced by policy fiat.

Hall's paper also highlights the importance of developing accurate models of traveler response to information for generating guidance and predicting its network-level impacts.

Finally, it argues strongly against attempting to manipulate ATIS messages (restricting or misrepresenting information) in an attempt to manipulate driver behavior towards some "social engineering" objective. Rather, he argues, ATIS should be viewed first as a service to the public, to improve their confidence and comfort in using the transportation system, and second as a means for steering traffic away from dis-equilibrium behavior and towards user optimal travel patterns that utilize alternate routes where feasible.

4.3 Conclusions from operational tests

With the possible exception of a few VMS-based ATIS in high volume corridors, operational experience to date with ATIS has been on too limited a scale and for too brief a time to be able to draw strong and broadly applicable conclusions regarding its network-level impacts.

Potential users have not had sufficient time to become aware of and comfortable with ATIS, and to integrate it into their travel decision-making processes. For this reason, among possibly others, the utilization of prototype deployments has generally been at low levels. The deployments themselves have usually been limited in capabilities, in time, and frequently also in geographic scope. Thus, the network-level impacts they have produced have often been small and difficult to measure, even when their impacts are more evident at the individual or (sometimes) corridor level.

[4] It is also possible that these studies made driver behavior assumptions that are not fully consistent with an equilibrium framework.

Some network-level ATIS evaluation studies have carried out their work in an indirect way. Rather than attempting to measure the impacts directly, they proceed by obtaining what (little) empirical data on impacts might be available, completing the data with default values and assumptions as required, and using the resulting data set as input to a traffic or economic evaluation model. The model then extrapolates the limited data to the full network level and computes the impacts.

This may be a reasonable approach until larger-scale ATIS deployments become common. It implies a well-defined and focused data collection effort tailored to producing data useful for such an approach, as well as network-level models capable of representing ATIS.

(Yim and Miller 2000) describe the evaluation of the two-year TravInfo field operational test by the California PATH program's Institute of Transportation Studies. TravInfo's goal was to broadly disseminate accurate, comprehensive, timely and reliable information on traffic conditions and multi-modal travel options to the public in the San Francisco Bay Area. To this end, it established a Traveler Advisory Telephone System (TATS), to which users could call for up-to-date information on travel conditions and options, as well as a web site displaying real-time traffic conditions. The evaluation considered the operational test from a number of viewpoints, including institutional, technological and user response.

With respect to user response, surveys showed that less than 10% of Bay Area households were even aware of TravInfo's existence or features and, of those who know about it, very few had actually tried it. Those who did use it, however, found the services to be useful for trip planning and reported high levels of satisfaction. TravInfo was able induce some users of radio and television traffic reports to switch, and also to to capture some people who had never before used radio or television reports. Roughly half of the TATS callers, and more than three-quarters of the website visitors reported altering their trips after obtaining information about their routes. By the end of the operational test, around 5 percent of users were asking about transit options and, of those, 90 percent reported using transit for their trip. Overall, however, the report concludes that TravInfo's impact on the transportation system was marginal.

(Lee 2000) presents a framework for the benefit-cost evaluation of Seattle, Washington's Internet-based freeway management system, called FLOW. Among other things, FLOW maintains a web site that displays a color-coded real-time prevailing traffic conditions on expressways and major arterials at the segment and lane level; data is updated every two minutes.

Very little data was available on the impacts of the FLOW system: mostly individual-level results from wave 7 of the PSRC survey and from an MMDI survey. Lee supplemented this data with default and assumed parameter values to estimate the economic benefits of different system impacts on different types of user. It expanded these to the total user population, and compared the resulting total benefits with a rough estimate of the system costs. The specific conclusions reached are perhaps not extremely reliable, given the many assumptions that were required to reach them. The interest and value of this paper is in providing an evaluation framework and methodology, and in indicating the types of data that would be required to carry out a more accurate evaluation.

The framework distinguishes five market segments by trip purpose, using purpose as a proxy for travel information needs and likely behavioral response. It also characterizes people as auto users, transit users and non-captives. The responses to information that are considered are: change mode; add trip; delete trip; change destination; change route, change departure time; change confidence level; and nothing.

Impacts of each of these response types are characterized as internal (to the traveler) or external (to others), and are evaluated; each response may entail a number of impacts of both categories.

Evaluation of internal impacts depends on whether the response is primarily motivated by consideration of travel times (as in route choice) or not. If so, travel time savings are estimated and converted to monetary equivalents. Otherwise, the paper suggests estimating the consumer's surplus change of the (possibly complex) response, using willingness to pay estimates from stated preference surveys.

External impacts are computed as the difference between marginal and average costs. Changes in modal VMT are used to compute external emissions costs, while changes in travel time are used to compute external congestion costs.

Computing benefits in this way, making assumptions about their change over time, and comparing the total benefits with rough estimates of system costs, Lee estimated that the benefit/cost ratio of the FLOW system was 2.0, with a range of uncertainty between 0.5 and 3.0.

(Wunderlich, Bunch et al. 2000) describe their model-based evaluation of the SmartTrek Seattle MMDI, involving ATIS (a variety of traffic information services) and ATMS (traffic signal coordination) measures in a freeway/arterial corridor north of the Seattle CBD. The evaluation focused on project impacts that are difficult to evaluate with direct field measurements because of their magnitude or geographic dispersion, or because of the presence of confounding factors.

The evaluation used both a conventional four-step transportation planning model (EMME/2) as well as a traffic simulator (INTEGRATION) that can use some planning model forecasts as inputs. The conventional model was used to identify regional-level impacts on travel demand patterns, while the simulation model was applied to identify ITS impacts under dynamic traffic conditions. Models were validated against corridor traffic counts and travel time measurements.

The simulation was applied to a series of scenarios representing combinations of traffic demand variations, weather conditions, and patterns of incidents. Each scenario has a weight, or probability of occurrence. The scenarios taken together comprise a representative year of system operation.

Different levels and combinations of ATIS and ATMS capabilities were tested against a baseline scenario. Evaluation measures included subarea and regional impact variables. Variables included delay reduction, throughput, traffic condition variability, VKT of travel, fuel consumption, pollutant emissions, mode shares, trip lengths and speeds.

5 MODELING THE NETWORK IMPACTS OF ATIS

This section discusses the modeling of ATIS and the prediction of its impacts on traffic flow patterns and conditions, within the framework of static and dynamic traffic network models.

Much of the literature reviewed in the preceding sections is concerned with the response of individual tripmakers to travel information; as the review indicated, the state of knowledge on this subject is still far from complete. Suppose, however, that very good models of individual tripmaker response to ATIS were actually available, so that it would be possible to accurately predict the departure time, destination, mode and/or route that a particular tripmaker would choose if he or she were to receive a particular set of information from an ATIS. Suppose, also, that many tripmakers received ATIS guidance. What then would be the overall effect on network flows and travel conditions resulting from the aggregate response of the individual tripmakers to the travel information that they received? Might the changes in flows and conditions be sufficiently large as to affect the travel information provided by the ATIS? And if so, how should the provided travel information account for tripmakers' responses? These questions are important if our knowledge of individual traveler response to information is to be usefully applied to improve network-level operating conditions, or to generate effective predictive guidance (see section 1.1).

Answering these questions requires a transportation network model that is capable of:

- adequately representing the technical and information characteristics of specific ATIS deployments as they affect tripmaker response and network impacts;
- accurately predicting tripmaker responses to received ATIS messages (as well as the travel decisions of those who do not receive ATIS messages); and
- translating predicted individual-level tripmaker behavior into the network-level travel flows and conditions that result from them.

Few if any current transportation network modeling packages can carry out these tasks in a direct fashion. Many current approaches to modeling ATIS in a network context utilize a two-phase approach. In the first phase, a transportation situation is analyzed using a conventional network model that does not represent ATIS services, and does not attempt to predict their effects on individual travel behavior or network-level traffic patterns and conditions. A post-processing adjustment of the conventional model outputs is then performed to account for the impacts of ATIS. This approach is typified by the IDAS (ITS Deployment Analysis System) software package. (IDAS is also able to analyze ITS technologies and services other than ATIS.)

The advantage of such approaches is that they do not require changes to currently used traffic modeling software and so can be applied immediately; IDAS, for example, can directly post-process outputs from a variety of commercially available software packages. On the other hand, the two-phase approach carries the risk of introducing inconsistencies between the procedures used in the conventional model and those applied in the post-processing stage. It would be preferable to accommodate ATIS fully and consistently within the framework of the network model system itself.

Some model systems have begun to do this. Both the DYNASMART-X system (developed at the University of Texas at Austin under the direction of Prof. Hani. Mahmassani) and the DynaMIT system (developed at the Massachusetts Institute of Technology under the direction of Prof. Moshe Ben-Akiva) incorporate travel information in some form in their network modeling. These systems are mesoscopic traffic simulators that build on the traditions of traffic network simulation modeling but add significantly to these traditions by incorporating features such as sophisticated driver choice modeling, dynamic OD matrix estimation, and others. Both are under continual development, and are ultimately intended for deployment and real-time use in an operational traffic information center. (A few commercially-available traffic simulation systems also incorporate information in their network models. Unfortunately, the suppliers of these systems are sometimes reluctant to provide detailed descriptions of the assumptions, methods models and algorithms incorporated in their software. For this reason these systems are not considered here.)

DYNASMART-X and DynaMIT incorporate many reasonable design decisions regarding the representation of travel information, the modeling of traveler response to information, and the incorporation of these ATIS aspects in a network model. However, it is fair to say that these design decisions were not made from the perspective of a fully general framework for incorporating travel information in network models. This comment is not intended as a criticism of either model. A general framework for network-level modeling of travel information did not exist at the time the software was being written. Moreover, the development of such a framework was not a high priority for either project; their primary concerns centered on the creation of very large yet reliable and efficient software systems for traffic simulation.

This situation can be compared, in some respects, to the evolution of conventional static network modeling approaches and software from early efforts to the present status. Traffic network modeling software developed during the late 1950s and throughout the 1960s was based on heuristics – reasonable-seeming methods that usually appeared to work efficiently and to give believeable results, but that could not be proven to be correct. This was so despite the fact that, in the 1950s, (Wardrop 1952) had clearly defined the notion of a traffic user equilibrium, and (Beckmann, McGuire et al. 1955) had formulated the equilibrium assignment problem as a well-posed optimization problem. However, it was not until the work of (LeBlanc, Morlok et al. 1975) that a rigorous and efficient algorithm for solving Beckmann's equivalent optimization problem was published, and software implementing provably correct solution methods became available. (As it turned out, some of the heuristics that had been developed were very similar to the rigorous solution method of LeBlanc et al., sometimes differing only in a single trivial step such as a line search.)

Thus, while the particular information modeling approaches adopted by projects such as DYNASMART-X and DynaMIT seem reasonable, there is benefit in attempting to develop a more general framework applicable to modeling network-level information impacts. The framework may suggest alternative modeling or solution approaches that, on examination, prove to be advantageous in some respect. At a minimum, the general perspective offered by the framework will provide a better appreciation and understanding of the particular design choices that were made during the development of existing modeling systems, and guidance for the development of new systems.

We propose here a high-level framework that might be suitable for this purpose. To this end, the following sections review the conventional transportation network modeling framework, highlight the difficulties encountered in applying this framework to ATIS modeling, and show how these difficulties can be resolved.

5.1 The conventional transportation network modeling framework

5.1.1 OVERVIEW

Entire books have been devoted to transportation network modeling (Sheffi 1985; Thomas 1991; Ortuzar and Willumsen 1996; Cascetta 2001), and it is not the objective here to duplicate them or to provide an exhaustive description of the current state of network modeling practice. Rather, the intent of this discussion is to briefly summarize the principal aspects of the conventional transportation network modeling framework, highlighting particular features that either lend themselves to or conflict with the needs of ATIS modeling.

Transportation network travel forecasting is an application in a network structure of the economic paradigm of supply-demand interaction leading to equilibrium. Given a description of network infrastructure and operational characteristics that supply transportation service, and of the land use and activity patterns from which travel demand is derived, transportation network modeling attempts to predict the demand flows and the network conditions that result from the supply-demand interaction over a particular analysis time period. Frequently, interest focuses on predicting the steady-state conditions that prevail over a period of time (e.g., a peak hour or peak period) that is long relative to the time scale of flow dynamics (e.g., the time taken by individual vehicle maneuvers such as lane changing, turning movements, or queuing); models that address this question are known as *static* transportation models, and software that implements them is widely available. More recently, interest has grown in replicating and predicting the variations in traffic flows and conditions at much finer time scales (for example minute by minute); models that address flow phenomena at this level of temporal detail are called *dynamic* transportation models. Most existing software packages for dynamic transportation modeling are research oriented, although commercial packages are beginning to appear.

Traditionally, network forecasting is carried out using the so-called *four-step process*, which consists of the following high-level operations:

- trip generation: determining the total number of trips produced by each origin traffic analysis zone (TAZ) and attracted to each destination TAZ in a study area over a particular analysis time period (for example, the peak hour);

- trip distribution: given the production and attraction totals computed during the trip generation step, determining the total number of trips traveling between each particular origin and destination zone pair (referred to as an OD pair);

- mode split: given the total number of trips between each OD pair, determining the number of trips made on each mode serving each OD pair. (This step may be omitted when analyzing situations where changes in mode shares are unlikely to be important); and

- assignment: given a fixed set of trips routing themselves through the modal (e.g. road or transit) networks from origins to destinations, determining the resulting volumes and travel conditions on network links (individual facilities).

There are numerous variations on the process. Frequently trips are distinguished by their purpose, and the generation, distribution and mode split steps are carried out separately for each distinct purpose (all trip purposes are aggregated in the assignment step, however). Some applications perform mode split before trip distribution. Others attempt to represent the choice of trip departure time, and so may shift travel demand from one analysis time period to another. Yet others incorporate a step that forecasts TAZ-level land use and activity patterns prior to trip generation.

Regardless of the details, however, the four-step process is fundamentally sequential, with each step proceeding to completion and providing its outputs as the inputs to the next step. Possible influences of later steps on earlier ones (in particular, the effects of travel conditions determined during traffic assignment on trip productions, attractions, distribution and/or mode split) are either ignored, or are accounted for via “feedback”: iterative execution of the entire sequence of steps using, in a given iteration, assignment step results from prior iterations (possibly with modification) as input to that iteration’s trip generation, distribution and/or mode split steps. The iterations are continued until some convergence criterion is met or a pre-specified computing effort is expended.

Note in passing that there is no intrinsic need to apply a sequential rather than a simultaneous approach to network forecasting. Indeed, unless “feedback” between the steps is performed correctly, use of a sequential approach may lead to incorrect model solutions (COMSIS 1996; Miller 2001). Numerous studies have shown how the trip generation, distribution and mode split steps (or some subset of these) can be integrated with the traffic assignment step in an extended model and correctly solved, but these results have not been widely applied in the transportation planning community, and few commercially available software packages for the static problem adopt this approach. In contrast, software for the dynamic problem tends to incorporate greater simultaneity between steps.

We will concentrate here on traffic assignment for road networks. Many of the major network effects of ATIS are captured in this step. Furthermore, of the four steps in the conventional model system, traffic assignment has arguably received the most thorough study, has achieved the greatest amount of consensus on valid approaches, and has been the most generally systematized in its implementation details (at least in static models). As a result, the discussion here can consider in some detail the standard methods for traffic assignment, and the modifications to them that are needed to incorporate ATIS in assignment. Because of the wide variety of approaches commonly applied to the other steps of the four-step process, a comparable degree of specificity is not possible with them. However, by focusing only on traffic assignment, we do not take account of the possible effects of ATIS on total tripmaking, destination choice, mode choice and trip departure time decisions, or consider network impacts of transit information systems. These effects can be handled by extending (but not fundamentally changing) the road network ATIS analysis framework that will be presented below.

5.1.2 STATIC TRAFFIC ASSIGNMENT

Given a demand for road travel between origin and destination zones over a particular time period, traffic assignment determines the link-level traffic volumes and conditions that result from the path choices made by the OD trips. In static traffic assignment, each OD pair’s demand is assumed to remain at the same level throughout the duration of the analysis time period, and the objective is to determine the resulting steady-state (i.e., time-invariant) link volumes and conditions that prevail during that period.

A traffic assignment model incorporates three basic components:

- a path choice model;
- a network loading procedure; and
- solution logic that initializes variables, invokes the path choice and network loading components with appropriate inputs, iterates as needed, and decides when the assignment process is complete.

Path choice models used in conventional traffic assignment models are based on economic utility theory: they assume that a tripmaker at the origin chooses, from among a set of considered paths to the destination, the path that is perceived to have maximum utility. However, different path choice models vary in their assumptions with regard to:

- how path utility is quantified. Usually path *dis*utility (to be minimized) is represented as travel time, travel cost, or some weighted combination of the two;
- how tripmakers perceive path utility; and
- how tripmakers identify the set of possible paths to consider.

Deterministic path choice models represent the path choice decision as if tripmakers have perfect information about network conditions when they are considering their path options at the origin. (For convenience of expression, the following discussion will frequently use travel time as a proxy for a more general definition of network conditions.) The available information is assumed to be completely accurate, correctly perceived, and to correspond exactly to the utility that tripmakers take into account when choosing a path (for example, travel time or cost minimization.) Tripmakers are assumed to consider all possible paths connecting their origin with their destination, and always to choose the path that has the highest utility (or least disutility) value. (If there are several such paths, any one may be chosen.) Standard minimum path algorithms are used for this purpose; they are able efficiently to determine the shortest (i.e., minimum disutility) path in a network without explicitly examining all possible paths.

In *random utility* path choice models, the utility of a path is represented as a random variable, usually specified as the sum of two terms: (i) a systematic (deterministic) utility similar to that used in deterministic models; and (ii) a zero-mean random disturbance having some given probability distribution. The disturbance term may represent imperfect perception of network conditions by the tripmakers; alternatively, it may derive from fundamental modeling limitations, such as the inability to capture in a model all of the personal considerations that enter into a tripmaker's path choice decision for a given trip. In either case, the disturbance reflects uncertainty with respect to the path choice decision made at the origin. Thus, although tripmakers are assumed to choose the path that provides maximum utility to them, the randomness of the utility specification prevents us from knowing with certainty which path that will be. Instead, we can only determine the *probability* that each of the different considered paths will be chosen: this can be computed from knowledge of each path's systematic utility, and the joint probability distribution of those paths' random disturbance terms.

It is neither behaviorally realistic nor computationally feasible to consider all possible paths in a network and compute a choice probability for each. Therefore, random utility path choice models generally apply some path pruning rule to select the paths that will be considered. These might involve a path efficiency criterion (requiring, for example, that each successive link on a path leads farther away from the origin), or generation of a small set of paths "on the fly" (i.e., in successive iterations of the assignment process), or *a priori* selection of a set of allowable paths.

Note, too, that some methods based on random utility path choice, such as the STOCH algorithm (Dial 1971), define the path choice set and determine the path choice probabilities in an implicit rather than an explicit fashion, by computing the probability of choosing each link exiting from each successive node in the set of efficient paths. However, the ultimate effect is the same as if the choice probabilities of efficient paths were being computed: given a specific path, its choice probability can be computed in a straightforward way.

After the path choice decisions are made, the traffic assignment process *loads* the corresponding trips on the selected path(s). Static network loading consists of:

- determining the part of the total OD flow that will use each of the available paths. The way in which this determination is made depends on whether a deterministic path choice model was used to identify a single path, or a random utility path choice model was used to compute the path choice probabilities of each of a set of paths. In *deterministic* loading models, all of the OD flow is loaded on the one selected path. In *stochastic* loading models each available path receives a part of the total flow, according to its probability of being chosen. (Again, in some loading procedures such as STOCH, the path flows are determined implicitly in terms of the flows exiting each node via each of the outgoing links;)

- propagating the path flow over each link on the path from origin to destination, and determining the resulting link flows. In static models, which consider only steady state conditions over the analysis period, the time required for flow to propagate from one link to the next, and the fluctuations in link volumes as flow traverses them, are not taken into account. The flow on each link is computed by accumulating the steady state flow contributed by each path that goes through the link; and

- updating link conditions based on the accumulated link flows. Any of a variety of relationships – generically known as link performance functions – may be used for this purpose. Link performance functions capture traffic congestion effects. Link condition updating is generally carried out after the flows from all OD pairs have been processed; it utilizes the final link volumes.

The path choices that are used in loading derive from a set of network conditions, yet these conditions may change as a result of the loading. Consequently, loading may invalidate the earlier path choice decisions: the path that was thought to be optimum in a deterministic model is no longer so, or the path choice probabilities computed from a random utility model no longer correspond to the new conditions.

All traffic assignment models attempt to enforce a system-level *consistency* requirement: after the assignment procedure determines tripmakers' path choices, loads them on the network and updates link conditions, the updated conditions should not cause a revision of the path choices. In *deterministic user equilibrium*, the paths that were thought to be minimum should remain so after the loading; and in *stochastic user equilibrium*, the path choice probabilities that were used for the loading should equal the probabilities that are determined from the updated conditions. The term *user equilibrium* originally relates to the definition by (Wardrop 1952) of the requirement that, in a deterministic path choice model, the paths that are chosen by travelers should all have equal travel times, and this time should be less than or equal to the travel time on any path not chosen. The term was later extended to cover user equilibrium in situations of stochastic choice as well. Although user equilibrium is not commonly referred to as a consistency requirement, in fact it is one. Inconsistency occurs when travelers choose a path thinking that they will encounter certain travel conditions, only to find that the conditions are actually different from what they expected; they would thus have cause to revise their path choice, so the original choices and conditions could not have been in equilibrium.

A fairly natural way of expressing an equilibrium condition in mathematical terms is as a fixed point equation. The following paragraphs explain this idea, which, as will be seen, generalizes in a fairly straightforward way to network modeling with ATIS.

Let f be a mathematical function mapping inputs x from a set X to outputs y in a set Y. We write $f: X \rightarrow Y$ to show the relationship of the function to its input and output domains and $y = f(x)$ to show the output value corresponding to a specific input value. Suppose that the input and output domains are the same; the function maps one value to another in the same domain. We say that a value x^* is a fixed point of the function f if $f(x^*) = x^*$; in other words, a value x^* is a fixed point of a function if, when it is supplied to the function as input, the function returns the same value as output: the function maps x^* to itself. A function may have zero, one, or multiple fixed points. To compute a fixed point of a function (if one exists), it suffices to solve the equation $f(x) - x = 0$ using any suitable root-finding method.

Suppose C denotes the set of possible link conditions, and P is denotes the set of possible path choice probabilities. A path choice model, regardless of how complex it may be, is simply a function D that takes inputs C and produces outputs P; we can write D: $C \rightarrow P$. Similarly, for a given fixed OD matrix, a network loader S, regardless of its details, is simply a function that takes inputs P and produces outputs C; we can write S: $P \rightarrow C$. The functions D and S are simply symbolic representations of whatever operations these two models carry out to transform their respective inputs into outputs.

For some path choice and loading models, the functions D and S can be written down mathematically if required; for others, the functions have no straightforward mathematical expression, and are defined only in terms of their algorithmic specification or software implementation. In either case, however, it is both possible and convenient in the discussion here to treat these models as "black boxes", ignoring their internal details and focusing only on their inputs, outputs and interconnections.

These two functions can be put together ("composed") in either of two ways. For example, if we start with a set of path choice probabilities p1, we can input these to the network loader to obtain the resulting set of traffic conditions c1 = S (p1). We can then use these output conditions to compute the resulting path choice probabilities p2 using the path choice model p2 = D (c1). If the "input" path choice probabilities p1 equal the "output" path choice probabilities p2 then, as was discussed above, p1 (or equivalently p2) represent user equilibrium path choices. The composition of the two functions in this order is written p2 = D°S (p1), which means: use p1 as input the to function S, then use the resulting outputs of S as inputs to the function D, calling D's outputs p2. From the discussion, it can be seen that equilibrium path choice probabilities are a fixed point of the composite function D°S: $P \rightarrow P$.[5] An equilibrium set of path choice probabilities p^* satisfies the fixed point equation $p^* = D°S\ (p^*)$.

Similarly, if we start with a set of conditions c1, we can use the path choice model to predict the corresponding path choice probabilities p1 = D (c1), and can then input these probabilities to the network loader to obtain a new set of conditions c2 = S (p1). If the "output" conditions c2 equal the "input" conditions c1 then, by reasoning similar to that presented above, c1 (equivalently c2) represent equilibrium network conditions, and these conditions are a fixed point of the composite

[5] This statement is exactly correct for stochastic user equilibrium, but must be slightly modified for deterministic user equilibrium. The issue is that, in the latter case, there may be several minimum paths at equilibrium, with some carrying traffic and others not. Thus, two equilibrium solutions may entail different path choice probabilities. This situation can be handled by a generalization of the fixed point concept to point-to-set maps. It is likely, however, that deterministic user equilibrium will not be widely used in ATIS modeling because it would imply, for example, that drivers will comply fully with route recommendations – a highly questionable assumption. Probabilistic models and stochastic user equilibrium (or its generalizations, discussed below) would seem to be more appropriate in ATIS modeling contexts. Thus, while the issue of non-unique paths in deterministic equilibrium must be carefully addressed in theoretical discussions, it is unlikely to be of much significance in practice.

function $S°D: C \rightarrow C$.[6] An equilibrium set of link conditions c^* satisfies the fixed point equation $c^* = S°D(c^*)$.

In the past, the principal application of fixed point equilibrium formulations has been to analyze theoretical properties of equilibrium flow patterns (see, for example, (Braess and Koch 1979)). They were rarely the basis for developing assignment algorithms, in part because of concerns over their computational efficiency and (for the path-based formulations) their computer memory requirements. Recently, however, Cantarella (1997), building on earlier work by Daganzo (1983), derived two separate fixed point formulations, in terms of link costs and link flows, for very general versions of the static traffic assignment problem, and showed that a simple algorithm called the MSA (discussed below) could be used to solve them.

Apart from its expression as a fixed point equation, the equilibrium consistency requirement has various other quantitative implications. In the case of deterministic assignment models, for example, one implication of the consistency requirement is that at equilibrium, for a given OD pair, conditions on all paths used by tripmakers are equal, and these conditions are better than or equal to those on any paths that are not used; as noted above, this is called Wardrop's ((Wardrop 1952)) principle. In the case of stochastic models, the quantitative implications require equality of the path choice probabilities or of link flows or conditions before and after loading. Consistency conditions such as these can be used as a basis for computing traffic network equilibrium or for checking a tentative solution.

Various mathematical formulations of traffic network equilibrium problem can be derived and their solutions shown to imply the equilibrium conditions. By far the most widely used of these are formulations as optimization problems. Half a century ago, Beckmann (1955) devised an optimization problem whose solution conditions imply the Wardrop principle for networks with a simple form of link performance function (i.e., one for which the traffic conditions on a link depend only on the total amount of traffic using that same link). Most current traffic network software packages compute deterministic equilibrium indirectly, by solving Beckmann's equivalent optimization problem using any of a variety of nonlinear optimization algorithms.

Similarly, Sheffi and Powell (1982), building on other earlier work by Daganzo (1979), formulated an optimization problem whose solution conditions imply the stochastic network equilibrium conditions. They identified a stochastic approximation algorithm that had been developed in the early 1950s, originally by Robbins and Monro and later by Blum, as being well-suited for solving their optimization problem, and named their version of this algorithm the method of successive averages (MSA). Most current software packages compute stochastic equilibrium by solving Sheffi and Powell's equivalent optimization problem (or special cases of

[6] The issue mentioned above for deterministic user equilibrium in path choice probability variables does not arise when condition variables are used. Thus, if DUE is used as a modeling principle, it will be easier to work with the S°D fixed point formulation.

it that result when particular path choice models are used), applying the MSA or other optimization algorithms for this purpose.

In both the deterministic and the stochastic cases, the optimization algorithms call on the path choice and network loading procedures to carry out particular computational subtasks required as part of the solution process. The algorithm implements the logic that determines the inputs that these procedures are invoked with, executes them, and ascertains if an equilibrium solution has been reached; if not, it continues with an additional iteration.

The MSA, for example, consists of the following steps when applied to solve for link cost traffic equilibrium conditions:

0:	c_0 = free-flow link costs; $k = 0$	*initialize*
1:	$p_k = D(c_k)$	*get path splits based on current link costs*
2:	$d_k = S(p_k)$	*get auxiliary link costs based on path splits*
3:	if convergence is achieved, stop	*if converged, c_k is the solution; else continue*
4:	$k = k+1$	*bump iteration counter*
5:	$c_k = c_{k-1} + (1/k) * (d_{k-1} - c_{k-1})$	*update link costs*
6:	go to step 1	*iterate*

The prototype version of this algorithm was originally developed to find the roots of functions whose evaluations return values tainted by noise; it can be mathematically proved to converge to a correct root value, under certain conditions on the involved function. Sheffi and Powell (1982) showed that these conditions are met when the method is applied to solve their equivalent SUE minimization problem, so for this problem it is a verifiably correct method.

The method has also been applied, however, to problems for which no convergence proof is available. It seems to work well on a wide range of problems, including even completely deterministic problems in which no noise is present. It does not require information about the mathematical properties of the involved functions (the only operation is function evaluation). It is very robust and simple to implement. Note, too, that one path split model could be replaced with another, or one network loader with another, without affecting the overall logic of the algorithm.

On the other hand, the lack of convergence proof for general applications is worrisome, and its computational efficiency is often disappointing: frequently, after achieving rapid progress in the first few iterations, it then slows down considerably and makes increasingly small improvements from one iteration to the next. The measurement and detection of convergence can also be delicate: it is not sufficient to compare two successive iterates (for example, c_k and c_{k+1} in the listing above) since, by the nature of the algorithm, they will tend to be closer and closer together as the computation progresses. A natural convergence measure, in the context of fixed point problems, is the distance between an estimate and its image through the map of interest; in the example given above, the distance (suitably defined) between c_k and $d_k = S°D (c_k)$. At a fixed point this distance would, of course, be 0. Even if some other convergence measure is used, verification of the fixed point condition after the algorithm has terminated is a direct way of checking the computed result when a rigorous convergence proof is lacking.[7]

Despite these issues, the MSA has been applied as a heuristic (i.e., a non-rigorous computational method that seems to work well in practice) to a wide variety of transportation problems, including the "feedback" operation in the four-step process, the dynamic network loading problem, and the dynamic traffic assignment problem, discussed below.

There are other mathematical formulations of the network equilibrium problem. One intensively studied formulation is as a variational inequality problem. The equivalent variational inequality formulation can be shown to imply network equilibrium conditions for very general link performance relationships, and so is more broadly applicable than the equivalent optimization formulation. In practice, however, the most commonly used link performance functions are of the simple form assumed by Beckmann and by Sheffi and Powell, so this added generality is not often needed in practical modeling work, and few commercial software packages compute traffic network equilibrium utilizing this approach.

5.1.3 DYNAMIC TRAFFIC ASSIGNMENT

5.1.3.1 Overview

Static network models have been used for over forty years in practical transportation network planning. Typical planning applications, concerned as they are with problems such as infrastructure project evaluation or meso-scale environmental quality analyses, do not usually need to consider the detailed dynamics of variations in traffic flows and conditions over short- and medium-term time frames, or transient traffic phenomena such as queue build-up and dissipation. More recently, however, there is increasing interest in being able to understand and predict dynamic (time-varying) features of traffic flow in networks, in part because of the importance of these capabilities for developing and operating real-time traffic management and information applications. Static models are not able to provide this level of temporal analysis detail. A different type of assignment model, called a dynamic traffic assignment (DTA) model, is required to represent the variations of traffic phenomena over time.

[7] This test is less straightforward when, as is the case in some stochastic user equilibrium models, function evaluations return values affected by noise; in such a case, the distance between c_k and $S°D (c_k)$ would generally be positive even if c_k were a fixed point.

This section considers DTA models as dynamic generalizations of the static models discussed above; this is the conventional usage of the term. Like static models, they incorporate a representation of tripmakers' path choice decision-making logic; include a network loader to propagate trips along the selected paths; and attempt to compute an equilibrium state involving consistency between network conditions and path choices. Furthermore, they assume that travelers have access to perfect information prior to setting out on a trip; they thus make a path choice decision at their origin and follow it through to their destination. (Clearly, such assumptions are not appropriate for modeling traveler information. Modifications to the conventional DTA modeling approach required to handle network-level impacts of traveler information are discussed in section 5.3.)

Although DTA models have similarities with static assignment models, there are also many important differences between the two. In dynamic models all variables are functions of time. Time may be represented as a continuous variable, or alternatively be discretized into finite duration time steps (ranging from fractions of a second to many minutes in length, depending on the desired temporal precision and the level of detail of the modeling relationships). OD demand rates may change from instant to instant, leading to surges or lulls in the amount of traffic entering the network. Similarly, the amount of OD traffic departing at each instant on the various available paths will change because of changing network conditions or randomness in driver behavior. Traffic may be represented as individual vehicles or vehicle "packets" (in simulation-based DTA models), or as flows or flow packets (in analytical models). As the traffic works its way along its chosen paths through the network, link traffic volumes and conditions will be affected by its passage and so also will change with time.

(In dynamic traffic models, the reference time for describing a time-varying phenomenon is, by convention, taken to be the time at which the phenomenon *begins*. For example, if we say that a particular link's traversal time is one minute at time t, we mean that vehicles that *enter* the link at time t will take one minute to traverse it.)

5.1.3.2 Path choice models

Because traffic conditions are functions of time in dynamic models, there is a greater variety of possible path choice model assumptions than is the case in static models. For example, the travel time required to traverse a path from origin to destination, as a function of the time of departure from the origin, can be defined in at least two different ways: as the sum of the traversal times of all the links on the path at the time of departure from the origin, or as the sum of each link's traversal time at the time the vehicle actually enters it. The former definition is called the *instantaneous* path time; it is based on a "snapshot" of the network conditions prevailing at the time of departure from the origin. The latter is called the *experienced* path time; it is the time that would actually be taken to traverse the path by a vehicle departing from the origin at that moment. (Note that the term *experienced* does not imply that the driver actually must travel on the path to learn its attributes.) Instantaneous conditions are easier to compute, but experienced conditions correlate better with what a trip encounters in its passage through the network.

The path choice component of a DTA model could involve either instantaneous or experienced travel time variables, depending on the data available with which to estimate the model, and the particular assumptions appropriate for the application. Nonetheless, in deterministic DTA models the path choice decision is represented as if drivers correctly perceive the (instantaneous or experienced) network conditions, while in stochastic DTA models the path (dis)utilities include a random disturbance term – just as in static assignment models. Like static models, DTA models assume that all information relevant to the path choice decision is available to tripmakers at their origins, prior to beginning their trips; similarly, DTA models assume that once tripmakers have chosen a path at the origin, they follow it unswervingly to the destination.

5.1.3.3 Dynamic network loading models

Recall that, in static models, the detailed changes in flows and conditions that occur as traffic propagates from link to link along its chosen path are not taken into account since these are transient rather than steady state phenomena. However, in dynamic models these variations are precisely the outputs of interest. The *dynamic network loader* allocates given fixed OD flow totals across the available paths in accordance with given path probabilities, then moves the resulting path flows from link to link along the OD path, simultaneously determining the resulting dynamic link flows and conditions. Network conditions determine the amount of advance per time step on each link and from link to link; conversely, network conditions are themselves determined by the dynamics of traffic propagating along its paths.

Dynamic network loading is considerably more complex than static network loading because it needs to track the progression of path flows from link to link over time, whereas static loading does not consider the progression but focuses only on its final result, i.e., the steady state link volumes and conditions. In static loading, given the path flows and the set of links making up each path, calculating the resulting link flows is a matter of simple addition; the flow on any link can even be calculated without considering other links. In a dynamic model, given the same kinds of information, determining the flow on a link *at a particular time* requires knowing the time-varying traffic conditions on upstream links, and these conditions are themselves affected by the flows on those links.

The dynamic network loading problem can be formulated in a natural way as a fixed point problem: dynamic link flows are determined by dynamic link traversal times, and the congestion these flows create must in turn result in the same dynamic traversal times. In simulation-based models, loading is frequently accomplished simply by "moving" simulated vehicles through the network in one time step, and then updating traffic conditions for the next step; a small time step is generally used to minimize the approximation errors introduced by time discretization.

Researchers have proposed a variety of ways of solving mathematical versions of the problem; some of them involve MSA-type procedures.

5.1.3.4 Dynamic user equilibrium

The task of a DTA model is to ensure that the network conditions that were the basis for associating trips with paths coincide with the conditions actually encountered by trips. Corresponding to the two definitions of path travel time discussed above are two distinct notions of dynamic equilibrium, termed *instantaneous* and *experienced* dynamic user equilibrium, respectively. If path choice based on minimum experienced time were assumed, for example, the model would attempt to ensure that a path that was thought to provide the least experienced time before loading did in fact do so after loading. DTA models attempt to determine a dynamic equilibrium in the sense that, according to the path choice assumptions incorporated in the model, and the network conditions that result when tripmakers pursue those choices over the network, no trip has an incentive at any time to change from the path it is following to some other path.

The first rigorous formulation of the dynamic traffic assignment problem is due to (Merchant and Nemhauser 1978); since then, mathematical approaches to formulate and solve dynamic traffic equilibrium problems have been the subject of active research. The mathematical analysis of such problems is highly complex (much more so than for the static problem), but through such analysis the basic characteristics of the problem and the properties of its solutions can be elucidated.

Much current work in the area involves the development and use of simulation-based dynamic traffic models, which lack strict mathematical rigor but allow an arbitrarily detailed representation of vehicle dynamics and driver behavior. Simulation models are generally felt to have the potential to provide the degree of modeling realism appropriate for practical, operational use in applications such as real-time network management or information provision.

5.1.3.5 Fixed point approach to dynamic traffic assignment

The dynamic traffic assignment problem also can be expressed as a fixed point problem involving the composition in either order of a path choice and a dynamic network loading map. In this case the underlying variables are time-dependent. We will write $p\{t\}$ and $c\{t\}$ for the path choice probabilities and link conditions, respectively, at a particular time t, and simply write p and c for the entire set of path choice probabilities and link conditions at all times (or time steps) over the considered analysis time period. The dynamic network loading model $S: P \rightarrow C$ takes as input the set of path choice probabilities P at all times over the analysis period, and outputs the corresponding link conditions C at all times. The path choice model $D: C \rightarrow P$ takes as input the set of link conditions C at all times over the analysis period, and outputs the corresponding set of path choice probabilities P at all times.

With these conventions, the DTA problem can be written as either of two fixed point problems that are formally identical to those that were discussed above for the static problem: $c = S \circ D(c)$ or $p = D \circ S(p)$. Because of the identical formal structure and other commonalities in the characteristics of the two problems, similar high-level fixed point solution algorithms can be applied to them; for example, versions of the MSA are commonly applied to solve DTA problems. However, the lower level details of the solution algorithms, such as the way that the solution variables are stored and accessed, or the operation of the network loader, are clearly quite different for the static and dynamic problems.

5.2 Difficulties of modeling ATIS in conventional DTA models

5.2.1 TERMINOLOGY

Some discussions of ATIS distinguish between prescriptive content (e.g., a route recommendation), which is referred to as *guidance*, and descriptive content (e.g., data about traffic conditions), which is referred to as *information*. This fine distinction will not usually be germane to the discussion here; indeed, as was noted in the literature review, messages that combine both descriptive and prescriptive content ("30 minute delays ahead, take alternative route XYZ") generally achieve the highest compliance rates by drivers. We will refer to any data provided to drivers by an ATIS as a *message*, and generally use the terms *information* and *guidance* interchangeably, unless otherwise noted. A specific message is characterized by its content and format, as well as its reception area and (in dynamic models) its time and duration of dissemination.

Three types of guidance can be distinguished, based on the type of information used to generate the guidance messages. *Fixed* guidance[8] provides travel-related information about things that rarely change. Examples of fixed guidance include guidebook-type information (locations and features of different attractions such as restaurants or museums) and basic way-finding directions that are not tied to actual traffic conditions. In contrast to this, reactive and predictive guidance base their messages on real-time measurements of actual traffic conditions over the network.

With *reactive* guidance, the guidance messages are based more or less directly on the real-time measurements of prevailing (instantaneous) traffic conditions; for example, a message might provide information about current traffic conditions on some link, or recommend a path that minimizes travel times as currently measured on links. With *predictive* (or *anticipatory*)

[8] The term *static* guidance is also used, but will be avoided here to prevent confusion with static traffic assignment models, a completely unrelated concept.

guidance, on the other hand, the real-time traffic measurements are combined with other data and used to make short- to medium-term forecasts of (experienced) traffic conditions throughout the network. These forecasts are then the basis of the guidance messages disseminated to drivers. For example, a message might provide information about what the traffic conditions will be on some link at the time in the future that the driver will actually arrive there, or suggest a path that minimizes the experienced travel time to the destination.

5.2.2 EXAMPLE

It was mentioned above that conventional dynamic and static assignment models assume that drivers have complete information on network traffic conditions relevant to their decision-making prior to their departure from the origin. Of course, if this assumption were at all realistic, there would be little need for ATIS! An ATIS provides travel-related messages to drivers precisely because their usual information basis for trip decision-making is imperfect, and drivers who exclusively rely on such “background” information might easily make sub-optimal travel decisions. Do such messages place tripmakers in the full information situation assumed in conventional traffic prediction models? Can a conventional traffic model be used to make the required traffic forecasts and to generate the guidance messages?

Consider a traffic network with a short-range source of traffic guidance – a variable message sign, for example, or a low-power infrared or microwave transmitter for in-vehicle receivers – located somewhere on it. The guidance source provides to nearby vehicles summary messages, based on real-time traffic conditions, about expected traffic conditions in the future.

Drivers leaving their origin do not have real-time traffic information available to them, so they must base their path choice on imperfect background information from some other source. Consider two identical drivers leaving at the same time from the same origin, going to the same destination and with the same background information. Consistent with a random utility path choice model, they may nonetheless decide to take different paths. Suppose that the path of one of the drivers does not go by the guidance source, while that of the other driver passes near the guidance source in the middle of the trip. The former driver will presumably follow to the destination the path selected at the origin. The latter receives guidance en route and interprets the summary message in some way. As a result, she may decide to switch to a different path, possibly to a different destination, conceivably even to a different mode. The aggregate tripmaking changes that result from many such drivers’ responses to the guidance messages may affect subsequent traffic conditions downstream of the information source, and possibly throughout the network. As a result of these changes, the network conditions are different from what was initially expected, and the guidance messages prove to have been incorrect.

5.2.3 DISCUSSION OF EXAMPLE

This simple example highlights most of the issues that must be addressed in modeling ATIS in a traffic assignment model.

First, guidance is provided because drivers without guidance have imperfect knowledge of the network conditions that are relevant to their real-time tripmaking decisions. It is therefore necessary to represent the no-guidance background information basis, and to model drivers' decision-making processes in this situation.

ATIS attempts to supplement drivers' no-guidance information basis with guidance messages. However, for technological or other reasons, the quality of the guidance information might itself be less than perfect. For example, guidance might not be available to all network users because its reception might require special equipment. Even to vehicles equipped to receive it, guidance might not be ubiquitous (available everywhere on the network) because its reception range might be limited to a relatively short distance from specific dissemination infrastructure such as VMS or infrared beacons. Constraints on communications bandwidth and human information processing abilities might reduce the level of detail and precision of the information that can be conveyed in the guidance messages: highly detailed and precise messages are unlikely to be available in many systems. Computation and communication delays might leave drivers with out-of-date guidance. The guidance itself might be inaccurately computed or perhaps corrupted during transmission.

It is readily conceivable that, for a given network and travel demand pattern, two guidance systems, differing in one or more of the above aspects – for example, the location and range of the guidance transmitters, or the quality and quantity of the information provided – may have very different impacts on network traffic conditions. It follows that realistic guidance system modeling must be able to represent the specific characteristics of the system and the information that it provides to drivers. This point is emphasized by (Dehoux and Toint 1991).

Having represented the characteristics of guidance information, realistic modeling must also accurately capture the diversity of possible driver responses to the disseminated messages. Some drivers may rely heavily on guidance information, interpreting it more or less effectively in their decision-making processes; others may choose to do the opposite of what the guidance suggests in an attempt to "avoid the crowd"; yet others may ignore it completely and follow their habitual choices. Modeling driver response to guidance messages will almost inevitably be more complex than traditional driver behavior and path choice modeling, in which questions of information format, content, availability and accuracy do not arise.

An additional issue in driver modeling is the response to ATIS messages in an en route situation. A pre-trip choice (with or without guidance information) is a commitment without immediate antecedent, whereas an en route decision may entail a reluctance to revisit or to revise a path choice that was made earlier in the trip. For this reason, an en route decision may exhibit some form of hysteresis or threshold effect. Conventional path choice modeling rarely needs to consider such effects.

The possibility of en route path switches also leads to a difference in the functionality that network loaders must provide in conventional and guidance-based traffic models. Conventional network loaders propagate flows on complete paths from origin to destination and determine the resulting network conditions. Guidance modeling requires a network loader that can re-route flows at en route locations in accordance with given path probabilities there, and can determine the effect of such path switches on link flows and conditions on the downstream sub-path between the switching location and the destination. Note that pre-trip guidance affecting the path choice decision made at the origin does not pose any new problem for a network loader – it must propagate traffic along complete paths, just as it must in conventional traffic models.

Finally, when guidance involves information about future traffic conditions, then network guidance modeling must ensure that the guidance is *consistent* – in other words, that drivers' reactions to guidance messages based on assumptions about the future do not invalidate those assumptions. Because the guidance is based on future conditions, some kind of forecasting model that takes account of guidance messages will likely be used. Consistency is simply the requirement that the inputs and outputs of this model do not contradict each other. It is a generalization of traffic equilibrium. Whereas the equilibrium concept applies to conventional traffic models in which information is implicit, consistency applies to guidance-based or similar models, in which information and drivers' reactions to it are explicitly taken into account.

Suppose, for example, that guidance is generated using a dynamic traffic network model based on experienced (i.e., predicted) travel times. The model might indicate impending congestion in a particular corridor. It would seem reasonable to disseminate guidance messages to warn drivers there or perhaps suggest an alternative route. However, if these messages are tested using the guidance model, drivers' reactions to them could well cause the congestion to shift, leaving the original corridor relatively uncongested and perhaps resulting in worse overall travel conditions. This shows that the guidance messages were not consistent – within the model, the forecast traffic conditions that were the basis of the guidance messages did not materialize after drivers received the messages and reacted to them.

In practice, of course, a guidance system's forecasts will have to be updated periodically to take into account the latest data collected from network sensors and to correct any inaccuracies that may have crept into the forecasts due to random disturbances or major disruptions such as incidents. Many systems have adopted a rolling horizon approach to handle this ongoing need to revise and update earlier predictions. The issue here is different: if the prediction and guidance generation procedure does not incorporate some way of including traveler response to the guidance in the forecasts themselves, the forecasts and guidance will be inconsistent and systematically biased. No amount of updating by a rolling horizon or other approach will be able completely to correct for this error.

The issues associated with consistency are perhaps most clear in the case, as above, of dynamic models with predictive guidance. However, they arise even in static guidance models. Note that there is no meaningful distinction in a static model between experienced and instantaneous travel times (or between predictive and reactive guidance) since, by assumption, the steady-state conditions accurately reflect both current and future flows and conditions over the analysis time frame. Guidance messages both reflect and affect the steady state, and so the consistency issue arises even here.

5.2.4 CONCLUSIONS

In summary, adequate incorporation of traveler information effects in network-level forecasting requires a traffic model that can:

- predict what driver behavior will be in the absence of guidance;
- represent guidance messages: their content, format, reception area, availability constraints and (in dynamic models) time and duration of dissemination;
- predict the effects of guidance on driver behavior, in the form of pre-trip decisions to choose a path and en route decisions to switch from one path to another;
- predict the traffic flows and conditions that ensue as a result of travelers' responses to guidance, particularly taking into account en route path switches; and
- ensure guidance consistency.

Note that if an ATIS actually provided perfect information (conforming to the assumptions in a conventional traffic model), and if travelers reacted to this information as assumed by conventional models, then application of a conventional traffic network model to generate guidance would be justified. The conditions predicted by the conventional model could be disseminated to drivers, and would be accurate, consistent guidance. However, if the information provided by the ATIS is less than perfect (in the various ways discussed above), then ATIS does not create a full information situation, and basing guidance on the predictions of a conventional model would be incorrect. A different kind of traffic network model is called for.

(Watling and van Vuren 1993) provides an excellent discussion of many detailed issues that arise in the network-level modeling of ATIS. The following section proposes an overall modeling framework that accommodates the inter-relationships between demand, supply and travel information in a network context.

5.3 A traffic network model framework for ATIS modeling

It is possible to identify and discuss the appropriate structure of a traffic network model suitable for ATIS analyses even if, as was seen in the literature review, the current state of art does not yet permit a definitive choice to be made regarding the most accurate model of traveler response to guidance or the best representation of the detailed characteristics of guidance messages. This is similar, in many ways, to the presentation of conventional equilibrium models given above, where it was possible to make fairly definite statements about the overall model structure and equilibration requirements, without considering in detail the specifics of particular path choice or loading models.

The following paragraphs present the major elements that are sufficient to include ATIS modeling in a traffic network model. This is not necessarily the most general framework, and its description is intuitive rather than precise, but the important ideas and features are present.

5.3.1 MODELING ELEMENTS

In addition to the usual links, nodes and centroids that define traffic networks in the conventional modeling approach, general ATIS modeling also requires the identification of decision points, a subset of the nodes at which tripmakers might choose or switch paths based on messages received in their vicinity. All origin zone centroids are decision points; some or all of the other nodes in the network might be as well, depending on the availability of en route guidance. A broadcast highway advisory radio system accessible throughout a metropolitan area might be represented by decision points at every node, whereas systems with limited reception range would be represented by decision points at nodes in the reception area only.

Decision points decompose origin-destination paths into subpaths between the nodes and the destinations. At each decision node, flow chooses a path to follow from that point towards the destination; it follows that path unless it encounters another decision node, in which case it may switch subpaths.

A basic framework for analyzing ATIS in a traffic assignment context is defined by three variables and three maps that relate them. The variables are path splits, link conditions and messages. The maps are the network loading map, the guidance map and the driver response map. Some of these modeling constructs are similar to, but not the same as, the corresponding constructs that were discussed above in the context of conventional traffic equilibrium modeling.

5.3.2 FRAMEWORK VARIABLES

Path splits P are the probability that drivers at a decision point will follow the various available paths leading from that decision point to their destination. They are a generalization of the path choice probabilities used in the conventional traffic network model, the difference being that path splits are defined not just at origins, but instead at any decision point.

Guidance information M is disseminated in the form of discrete units called messages. The representation of a message involves its content and format, the location (decision point) where it is disseminated and, in dynamic models, the time and duration of its dissemination. Further distinction may be made between different classes of messages available to different classes of users (e.g., radio messages that require special equipment to receive and decode). It is assumed that vehicles on links immediately upstream of a decision node can receive messages disseminated there (if they are equipped to receive them.)

Link conditions C are identical to their definition in the conventional network model.

5.3.3 FRAMEWORK MAPS

The network loading map S determines the link conditions that result from the movement of the exogenous OD demands over the network in accordance with given path splits P. As discussed above, network loading maps for ATIS applications must be able to handle en route path switches, something that conventional loaders are not designed to do. At the time of this writing, there does not appear to be any rigorous analysis of network loaders of this type published in the technical literature. However, there is no particular difficulty in understanding, in algorithmic or software terms, how such a loader would operate; indeed, packages such as DYNASMART-X and DynaMIT incorporate dynamic loading software modules with re-routing capabilities.

Operations that a loader carries out in a conventional traffic model for traffic departing from the origin must be carried out in ATIS models for traffic leaving any decision point. Whether at an origin or an en route decision point, the loader needs to allocate outgoing traffic among the available paths or subpaths in accordance with the path splits, as well as propagate vehicles downstream from the decision points, do any necessary bookkeeping, and determine the travel conditions that result.

In static models, this can be accomplished by applying a recursive loading procedure that, at each decision point beginning with the origin, splits path flows and propagates the resulting subpath flows along links until another decision point is reached. At each decision point the procedure is recursively re-invoked. A recursion terminates when the destination is reached.

In dynamic models based on vehicle simulation, this can be accomplished by determining, for each vehicle about to leave a link, whether the link's end-node is a decision point; if so, the vehicle selects of the available downstream subpaths to the destination in accordance with the path splits at that node. For dynamic analytical models, the variety of network loading methods makes it difficult to generalize; however, some of these methods group flows into packets which are moved rather like vehicles in a simulation system, so the method proposed for simulation models would apply, with appropriate modifications, to these models as well.

The guidance map G determines the messages that will be disseminated in response to a given set of network conditions. It can be thought of as representing the message selection strategy applied by the traffic information center being modeled. Recall that, by definition, a message is location- (and, in dynamic models, time-) specific. Thus, the guidance map determines not just what messages to display, but where (and when, and for how long) to display them. Although the link condition predictions output by the loading map are complete (i.e., they cover all network locations and, in dynamic networks, all times), there is no requirement that the generated guidance messages convey guidance information or recommendations having a comparable degree of detail, precision or network coverage. A set of detailed condition predictions from the network loading map might be summarized in messages such as "expect congestion ahead for next 30 minutes" displayed at one or two locations; a complete minimum path calculation might be summarized as "turn left at next traffic signal", with further routing information provided at subsequent decision points. There is no corresponding map in conventional traffic models.

The driver response map D predicts the path splits that result from drivers' responses to a given set of guidance messages. The map is a generalization of the path choice component of conventional traffic assignment models. In conventional models, the path choice model relates path splits at trip origins to path attributes, which are in turn are directly derived from link conditions; under the full information assumption, these are assumed to be known to drivers. In guidance models, on the other hand, path splits at decision points (origins or otherwise) are related to the guidance messages that are available there, and these messages are indirectly derived from network conditions via the guidance map. The map encapsulates the effect of guidance messages on path splits. In reality, of course, drivers may base their route choice decisions on a wide variety of other factors, including their general knowledge of traffic conditions, their prior experience (if any) with the guidance system, etc. It is not intended to neglect these other influences; rather, it is assumed that they have been subsumed in the driver response map, leaving a direct relationship between path splits and guidance messages only.

It is worth emphasizing the generality of the framework, and the way in which its various components are interrelated. For example, we have frequently used travel time as an example of a condition variable, and travel time-based descriptive or prescriptive information as an example of guidance messages. Actually the choice of condition variable can be essentially arbitrary.[9] What is important is that, given path splits, the network loading map must be able to compute whatever particular condition variables have been chosen. Similarly, the format and content of the message variables is essentially arbitrary. What is important is that the guidance map must be able to generate the appropriate messages for any possible condition variable values that may be output by the loader. Finally, the guidance map must be able to predict the path splits that result from any set of messages that may be generated by the guidance system, but the way in which it does this can be arbitrary.

5.3.4 COMPOSITE MAP FORMULATIONS OF GUIDANCE CONSISTENCY

The above considerations lead naturally to the definition of composite maps that combine the network loading, guidance and driver response maps in different sequences. Each composite map takes input in the form of a value of one of the variables discussed above, and transforms it into a (possibly) different value of the same variable. In fact there are three such composite maps:

- a composite map $D°G°S: P \to P$ from the domain of path splits into itself, which starts with path splits, forecasts the corresponding network conditions, determines an appropriate set of guidance messages, which are disseminated to drivers and cause them to respond in some way, leading to a new set of path splits;

- a composite map $S°D°G: C \to C$ from the domain of link conditions into itself. The map begins with a set of link conditions and determines the messages which the ATIS disseminates about them; these are communicated to drivers, who respond and possibly change the path splits; the flows propagating over the network in accordance with these changed path splits then lead to a new set of conditions;

- a composite map $G°S°D: M \to M$ from the domain of guidance messages into itself. Here the map begins with a set of messages, predicts the resulting path splits, forecasts the network conditions that ensue from these, then determines a new set of messages appropriate for these conditions.

In operational terms, evaluating one of these composite maps corresponds to executing one iteration of an ATIS network forecasting model that invokes the component maps in the indicated order of composition. The input to the model is an assumption about the value of one of the modeling variables (path splits, conditions or messages); its output is a prediction of a

---

[9] Actually, in dynamic models travel time *must* be one of the variables output by the loader because of its role in establishing traffic dynamics. However, there is no requirement that the guidance map take account of travel time in generating messages, for example (“incident ahead, take alternate route”).

possibly different value of the same variable. This can be written using a "functional" notation {i.e., function (input) = output} as:

model (assumptions) = predictions

Recall that guidance generated by a model is said to be consistent when the assumptions used as the basis for generating it prove to be verified, within the logic of the predictive model, after drivers receive the guidance and react to it. In terms of the composite maps, consistency means that a map's output predictions coincide with the input assumptions. Again, this can be written as:

model (assumptions) = predictions = assumptions

For the composite path split map, guidance is consistent if the forecast path splits coincide with the splits that were assumed at the start. For the composite network condition map, guidance is consistent if the initial network conditions used for the guidance determination coincide with those that are predicted to result after the guidance is disseminated. For the composite message map, guidance is consistent if the resulting messages coincide with the initially-assumed set of messages. Under mild conditions, solving any one of these problems is equivalent to solving any of the others.[10] There is not yet enough experience acquired with ATIS network models involving the different composite formulations to draw definite conclusions regarding their advantages and disadvantages, either theoretical or computational.

It can be seen that guidance consistency corresponds to a fixed point of a composite map that combines the relevant problem relationships. By solving one of the fixed point problems, a consistent value for the corresponding variable is determined, and from that value the solution values of the other variables can also be found. For example, if a fixed point $c^* = S \circ D \circ G\,(c^*)$ of the composite condition map is found, the resulting condition values account for the effects of the guidance messages on driver behavior, and the impacts of this behavior on network conditions. The consistent messages m^* can then be found by evaluating the guidance map using the fixed point conditions: $m^* = G(c^*)$. The driver responses (i.e. path splits) p^* to these messages can then be found via the driver response map $p^* = D(m^*)$.

Unlike conventional static and dynamic equilibrium problems, for which a variety of significantly different formulations are available, to the best of our knowledge the only approach currently available for general ATIS network modeling problems – involving a realistic representation of the guidance system and fully accounting for consistency – is via a fixed point formulation.

---

[10] Again, formulating a problem having a deterministic driver response map in terms of path splits may lead to a more complex situation because in this case path splits may not be uniquely defined at consistency. This problem does not occur if the composite condition or composite message maps are used.

A number of researchers have proposed ATIS modeling approaches similar in spirit to that described here. (Rilett and van Aerde 1991b) argued for the importance of providing routing information based on anticipated travel times. (Kaufman, Smith et al. 1991) was an early effort that treated guidance generation as a dynamic traffic assignment problem, but proposed a fixed point approach to formulate and solve it. Aspects of this approach to guidance via the DTA problem were pursued and analyzed more rigorously in (Kaufman, Smith et al. 1998). (Kaysi, Ben-Akiva et al. 1993) considered a more general notion of guidance, clearly defined the notion of guidance consistency, and evoked the possibility of analyzing it using fixed points, but did not pursue or formalize this idea. (Engelson 1997) also recognized the importance of predictive consistency, and its fixed point interpretation, in DTA-based guidance generation. (Bovy and van der Zijpp 1999) considered a general guidance system and analyzed it with a particular fixed point formulation. The framework proposed here is a generalization of these prior approaches. Its elaboration in a dynamic network context is presented in (Bottom 2000).

5.3.5 RELATIONSHIP TO EQUILIBRIUM MODELS

A conventional full information equilibrium model assumes that the path choice decision is made at the origin and is not reconsidered en route. Drivers are assumed to have accurate perceptions of the attributes of alternative paths and to choose a path that maximizes their perceived utility (although this choice may nonetheless seem random to a modeler who is not fully aware of the driver's decision situation). The loader propagates traffic along these paths from origin to destination, and determines the corresponding traffic conditions.

In terms of the guidance modeling framework proposed here, the only decision points in equilibrium models are at the origin. The network loader does not need to handle en route path switch situations. The guidance map (transforming network conditions into messages) is a kind of identity map I: the messages perfectly convey the exact network conditions. The driver response map reacts to these fully-informative messages in the same way that drivers are assumed to react to full information on conditions in the conventional model.

In this situation the composite condition map $S°D°I : C \rightarrow C$ and the composite message map $I°S°D : M \rightarrow M$ become equivalent. Only two distinct maps remain: the composite path split map $D(°I)°S : P \rightarrow P$, which is the same as $D°S : P \rightarrow P$; and the composite condition map $S°D(°I) : C \rightarrow C$, which is the same as $S°D : C \rightarrow C$. But these are the same as the two composite maps that were considered above in the discussion of fixed point formulations of conventional models. Consistent guidance, in this context, involves a fixed point of one of these two composite maps, which, as was seen above, is equilibrium.

To summarize, the conventional full information equilibrium model can be viewed as a special case of a guidance model in which the guidance information is perfect.

5.3.6 SOLVING ATIS NETWORK MODELS

As has just been seen, solving an ATIS network problem to obtain consistent guidance and its impacts can be accomplished by finding a fixed point of the composite map that is chosen to represent the problem. The basic approach for computing such a fixed point is no different in principle from that used to compute fixed points of conventional equilibrium problems. The general ideas are sketched out, from the viewpoint of developing software to address the problem, in the following paragraphs.

Fundamental choices about the type of model (i.e. static, simulation-based dynamic, analytical dynamic) have to be made at the beginning and, conditional on the choice, capabilities provided to support the basic operations (network creation and manipulation; data input, storage, access and output; time functions in dynamic models; etc.) that the model type requires. Many of these functions are fairly generic within each type of model. It is possible that portions of code prepared for conventional equilibrium models could be reused for these purposes (this would depend, of course, on implementation details).

Consensus has not yet been reached on some fairly basic issues in ATIS network modeling, such as the most appropriate representation of guidance messages or the form and specification of the driver response or message generation maps. Consequently, it may be most straightforward and efficient to tailor a software implementation to the particular problem at hand, rather than attempt to provide capabilities to handle these issues in a very general way.

Clearly, software to implement each of the component maps (i.e., network loader, driver response and guidance message generation) will need to be prepared; each map should be implemented as a distinct function (in the programming sense), accepting and returning the appropriate types of argument. The composite map constructed from these three components should be as efficient as possible, since its evaluation will be the bottleneck in the solution algorithm.

The MSA can be applied to compute the fixed point of the composite map chosen for the guidance problem, just as it is for conventional equilibrium problems. Of course, the same caveats apply to the two problems. In particular, no general result guarantees that the MSA applied to this problem will converge to a fixed point. Despite this, the observed performance of the algorithm is frequently satisfactory and, in problems without significant noise, the correctness of the computed solution can easily be checked by verifying the fixed point property.

As an example, the listing below shows the MSA logic applied to the composite link condition formulation of an ATIS network model. The algorithm evaluates the link condition composite map $d_k = S°D°G\,(c_k)$ in steps 1—3. This evaluation would be accomplished in the software by calling in succession the functions implementing the guidance map, the driver response map and the network loader. The close structural resemblance of this algorithm to the one presented above for the composite link condition equilibrium formulation is evident.

0:	c_0 = free-flow link costs; k = 0	*initialize*
1:	$m_k = G(c_k)$	*get messages based on current link costs*
2	$p_k = D(m_k)$	*get path splits based on messages*
3:	$d_k = S(p_k)$	*get auxiliary link costs based on path splits*
4:	if convergence is achieved, stop	*if converged, c_k is the solution; else continue*
5:	$k = k+1$	*bump iteration counter*
6:	$c_k = c_{k-1} + (1/k) * (d_{k-1} - c_{k-1})$	*update link costs*
7:	go to step 1	*iterate*

The often slow convergence of the MSA was mentioned above. This can be a particular problem in dynamic models because of the number of elements (dimension of problem variable multiplied by number of time steps) that must be adjusted to reach a fixed point, and because dynamic network loading is typically a computationally intensive procedure. The problem is even more acute in dynamic ATIS network models when they are intended to generate guidance in real time.

A number of methods to improve the convergence rate of the MSA (and similar algorithms) have been proposed over the past decade. While some of these are heuristics, an algorithm by Polyak (1990) (see also Polyak and Juditsky 1992)) is rigorously applicable whenever the MSA is, and can be shown to have optimal convergence properties in a certain sense. Polyak's algorithm is easy to implement and represents a very minor additional computational effort beyond the MSA algorithm. It also seems to improve the performance of the MSA even in applications where the MSA is not provably convergent. Methods such as Polyak's hold considerable promise for improving the solution speed of ATIS network models in both planning and real-time applications.

(Bottom 2000) describes software that implements fixed point approaches for dynamic network traffic modeling with ATIS, and tests the MSA and Polyak algorithms as solution methods. It was found that the Polyak algorithm could outperform the MSA (in terms of the number of iterations required to attain a certain degree of convergence) by factors of four or more. Its further application to fixed point formulations of traffic network problems would appear to be very promising.